Fear and Loathing in the Housing Market: Evidence from Search Query Data[*]

Marcelle Chauvet[†]
University of California, Riverside

Stuart Gabriel[‡]
University of California, Los Angeles

Chandler Lutz[§]
Copenhagen Business School
First Version: September 25, 2011

September 16, 2013

Abstract

We use Google search data to develop a broad-based and real-time index of housing distress. Unlike established indicators, this new Housing Distress Index (HDI) directly reflects sentiment revealed through search queries. Research findings indicate that the HDI provides important new insights in the determination of subprime mortgage credit-default swaps, the VIX, foreclosures, and national or city-specific housing returns. Further, the effects of housing distress on returns are asymmetric and stronger during times of crisis. Overall, results suggest that real estate and related markets are highly sensitive to vulnerabilities in household finances as captured by our measure of housing distress.

JEL Classification: G02, R30, G12;

Keywords: Housing Crises, Housing Distress, Housing Fear

The financial concerns of homeowners are of paramount importance to the US economy. As of 2011, there were over 45 million home mortgages outstanding in the United States valued at nearly 10 trillion dollars.[1] As evidenced during the 2000s housing and financial crisis, higher incidences of homeowner mortgage distress can dampen future house prices, induce episodes of pessimism among consumers and business investors, and wreak havoc on the macroeconomy or financial markets. Indeed, the crisis was accompanied by high levels of market volatility, reduced economic output, and widespread investor fear of financial meltdown. At the time, negative sentiment was captured by a broad array of market indices, notably including a popular equity market investor fear gauge known as the VIX (Whaley (2000)). However, neither the VIX nor other risk proxies provided much insight as regards sentiment in housing markets. Further, the few available measures of housing market sentiment were characterized by limited sampling, low frequency, and a lack of predictive power. The paucity of information was indeed striking, given the leading role of housing in the global downturn.

In this paper, we aim to fill this gap via the development and test of a new measure of housing market investor sentiment. We seek a broad-based, real-time gauge capable of capturing the palpable swings in sentiment that pervaded the housing and mortgage markets over recent years. Akin to the role of the VIX in equity markets, the new index should provide a proxy for investor fear in housing and mortgage markets while at the same time providing new insights regarding returns and related outcomes in those markets. In developing and testing such an index, we apply new, aggregate internet search query information from Google. We aggregate Google search queries for terms like "foreclosure help" and "government mortgage help" to compile our proxy measure of financial distress and fear in housing and related markets (henceforth, HDI).

Using Google data, we collect sensitive information directly from individuals conducting an internet search on issues of housing foreclosure and mortgage default.[2] All search terms utilized in the index relate to housing market financial distress. When a

---

[1]Mortgage statistics are from the US Census Bureau ACS Community Survey 2011 via the FactFinder website and debt totals are from the Federal Reserve Statistical Release z.1. December 6, 2012, P. 9, Table D.3.

[2]Conti and Sobiesk (2007) find that users are forthcoming when using internet search engines.

user enters a search term into Google such as "mortgage foreclosure help," he *divulges* his concern about mortgage failure or foreclosure. This makes our housing distress index unique compared to other indices in the housing literature. Other housing variables arise from surveys or market outcomes.[3] Surveys by organizations such as the University of Michigan or The Conference Board do not ask sensitive questions about mortgage distress and it is not clear that respondents would answer such questions truthfully.[4] Furthermore, proxies such as foreclosures are the result of equilibrium outcomes whereas other indicators like the VIX may reflect economic influences unrelated to housing markets. Our housing market distress index is related to these and other housing measures, but responds differently to market events. Thus, our index captures a new dimension of household behavior not previously observed in the literature.

Over our study period, the HDI is strongly inversely correlated with house prices.[5] Figure 1 shows a plot of the HDI versus the Case-Shiller 20 City National Housing Price Index. The HDI stays low during the housing bubble years but then jumps in 2007 and peaks at the height of the crisis in March 2009. After the crisis, the HDI falls but does not return to its pre-2007 lows.

The HDI is also strongly related to survey measures of housing sentiment as published by the University of Michigan. The correlation between the HDI and survey responses that cite that "now is a bad time to buy a home due to an uncertain future" (henceforth, *UncertainFuture*) is 0.74. The high correlation with negative housing sentiment lends credence to the interpretation of our index as a fear gauge for the housing market. Yet *UncertainFuture* and the HDI behave differently over our sample period. The HDI rises at the onset of the crisis in 2007 and then falls as market turmoil subsides. In comparison, *UncertainFuture* jumps only after the Lehman Crisis and remains at elevated levels throughout the remainder of the sample. Thus, the search query data captures a different element of household behavior than *UncertainFuture*. Increases in the HDI also lead and

---

[3]Da, Engelberg, and Gao (2012) make a similar point when they develop a stock market sentiment index using Google data.

[4]Singer (2002) contends that survey respondents are less likely to truthfully answer sensitive questions.

[5]The correlation coefficient between the first difference of the HDI and returns on the Case-Shiller 20 City National Housing Price Index is -0.221.

predict increases in negative housing sentiment.

In addition to our housing distress measure, we build an alternative index that is orthogonalized to key macroeconomic and housing indicators (henceforth, HDI$^{\perp}$). Through HDI$^{\perp}$, we can examine the relationship between the HDI and systematic macro or housing variables. We find that the macro and housing proxies explain little of the variation in HDI and that our unadjusted and orthogonalized indices are highly correlated.[6] Thus, our results are largely robust to the use of alternative housing distress measures.

Using the HDI, we study the predictive effects of housing market distress on an index that tracks the costs of subprime mortgage credit-default swaps, foreclosures and delinquencies, the VIX index, national housing returns, and housing returns over the geographical cross-section.[7]

First, we compare the HDI to the Markit ABX indices. The ABX indices are market based tools that track the cost of credit-default insurance on subprime mortgage-backed securities and are closely watched on Wall Street.[8] The results indicate that changes in the HDI predict the returns on the ABX indices. As the ABX indices are a liquid measure traded frequently by practitioners and institutional investors, our findings suggest, via an efficient market hypothesis argument, that household signals of distress as recorded by Google searches and the HDI were initially unbeknownst to and later discovered by market participants. Thus, our index appears to capture a new dimension of household behavior not previously observed in the literature. Overall, our results regarding the HDI and subprime credit-default swaps are noteworthy as the ABX indices acted as a leading indicator of the widespread turmoil that permeated through virtually all markets during the recent crisis (Longstaff (2010)).

We also examine the HDI in relation to foreclosure outcomes and delinquencies. Even

---

[6] The adjusted R^2 of the regression of the first difference of the HDI on the set of macro and housing variables is 0.15; suggesting that systematic factors explain little of the behavior in the HDI. The correlation coefficient between dHDI and dHDI$^{\perp}$ is 0.921.

[7] As the measurement of internet search query behavior is a relatively recent innovation, we study the predictive effects of the HDI in a Granger causality type framework. Overall, given the length of the sample, the strong predictive effects that we document throughout this study suggest that the HDI captures a substantial portion of the information content found in the forecasted variables.

[8] *The Wall Street Journal.* June 21, 2007. "Index With Odd Name Has Wall Street Glued; Morning ABX.HE Dose"

though the HDI is highly correlated with foreclosures, the two series differed dramatically throughout the sample period. During the crisis and its aftermath, the path of foreclosures was dramatically altered by legal challenges and public programs designed to aid defaulted borrowers and slow evictions. These programs included Project Lifeline and the California Foreclosure Prevention Act.[9] Further, foreclosures fell markedly in the US in 2010 and 2011 after the discovery of widespread lender abuse of foreclosure practices (so-called "robo-signing" scandal). In contrast, despite some temporary abeyance of foreclosure activity, the HDI continued along its previous trend. This implies that although the government programs managed to temporarily limit foreclosures, they had less of an effect on behavioral measures of financial distress. Moreover, we find that the HDI predicts foreclosures, prime delinquencies, and subprime delinquencies. These results are robust to the inclusion of controls for the various government housing assistance programs initiated over the sample period. Thus, households appear to signal their concerns regarding their financial distress via Google Searches before they become delinquent or enter foreclosure.

As previously noted, the VIX index is commonly known as a fear gauge among investors in the stock market. The HDI and the VIX are highly correlated, but the two measures capture investor sentiment in different markets. The VIX reacts more to stock market events such as the Lehman crisis, whereas the HDI is more highly correlated with housing indicators and survey-based housing sentiment. Also, increases in the HDI predict increases in the VIX. This result is not surprising since the housing bust led to turmoil in the stock market.

With regard to prices, we find that an increase in the HDI predicts a drop in national housing returns. This effect is more pronounced during times of crisis. Hence, the relationship between the HDI and housing returns is asymmetric. Like the VIX, the HDI is more a barometer of fear associated with the implosion in housing than a measure of agent exuberance during an upturn in the market. Graphically, we can see this in figure

---

[9]Project Lifeline was a collaboration between the Treasury Department and major mortgage banks. Through Project Lifeline, major mortgage bankers agreed to halt foreclosures for 30 days and then work with households to stave off foreclosures. In the California Foreclosure Prevention Act, a 90 day moratorium was placed on foreclosures. Information about the California Foreclosure Prevention Act can be found at `http://www.yourhome.ca.gov/cfpa_faq.pdf`.

1. The asymmetric relationship between the HDI and the national housing returns is qualitatively similar to the asymmetric relationship between the VIX index and stock returns. In the case of the latter, Whaley (2000) and Szado (2009) find that increases in the VIX are highly negatively correlated with large drops in stock prices.

Our findings regarding the importance of housing distress during times of market turmoil match the theoretical predictions of models that use leverage to explain asset price fluctuations. For example, in the housing literature, Stein (1995) constructs a framework where agents use leverage to purchase homes and finds that a deterioration in the agents' housing-related financial situation inhibits them from using the equity in their current house to purchase a new home.[10] This then creates a drop in housing demand and a decrease in prices.[11] Thus, Stein's framework provides a key interpretation of the predictive relationship between the HDI and house price returns: higher incidences of housing distress signal a worsening of housing-related finances, which, in turn, impairs the ability of agents to purchase new homes using available equity and leads to a decrease in demand and future prices.

We also examine the predictive effects of the HDI on housing returns across FHFA divisions and Case-Shiller cities.[12] First, we employ a panel data model and find that increases in the HDI predict a drop in future returns over the entire cross-section. Then, to accommodate potential heterogeneity in the relationship between our index and returns across local housing markets, we consider aggregated portfolios based on housing volatility or momentum in conjunction with our housing distress measure. The results indicate that an increase in the HDI leads to a larger drop in future housing returns for the most volatile or poorest performing housing markets. In other words, the predictive effects of the HDI over the cross-section are asymmetric and are most pronounced for more speculative housing markets and those in distress. In a similar fashion, Ang et al. (2006) find an

---

[10]For applications in other fields, see Shleifer and Vishny (1992) and Kiyotaki and Moore (1997).

[11]Lamont and Stein (1999) consider the leverage-to-value ratio over the cross-section and find that house price returns in cities with a high portion of highly leveraged households react more to income shocks.

[12]Over the cross-section, the FHFA reports monthly house prices based on the nine geographical divisions used by the US Census Bureau. The Case-Shiller data covers the largest 20 housing markets in the country.

asymmetric relationship between the VIX index and the cross-section of stock returns as stocks that are more sensitive to innovations of the VIX have lower future returns.

Use of Google search data to develop a housing market distress index has a number of distinct advantages. First, as noted above, our data allow us to directly observe the relative frequency of Google internet searches on issues of housing distress and thus to infer the extent of agent negative sentiment. Second, the Google search data aggregate queries from millions of households across the United States. Accordingly, we have access to a substantially larger sample than the typical survey. Third, Google data are available in real time (Choi and Varian (2012)). In comparison, traditional housing indices are available only with some lag. Fourth, the search data used to construct the HDI are free of charge and easy to obtain, whereas some other housing datasets, including foreclosures and the ABX indices, must be purchased. In sum, our index is useful since it is broad-based, directly captures housing market sentiment, predicts common housing variables, and is available in real time.

Although the HDI is to the best of our knowledge the first application of Google data to measure housing distress, other papers have applied search query data in other contexts. Ginsberg et al. (2009) used search query data to predict the spread of influenza in the United States. In the finance literature, Da, Engelberg, and Gao (2011) used search data to develop a new measure of stock market attention, whereas Da, Engelberg, and Gao (2012) applied search query data to capture consumer sentiment. While that latter paper provides insights into stock market sentiment, Da, Engelberg, and Gao (2012) do not focus on housing nor do they use any housing related terms in their index.[13] Mondria, Wu, and Zhang (2010) use Google data to study international home bias and attention allocation. With regard to housing, Beracha and Wintoki (2012) predict local housing prices using the search term "real estate" in conjunction with the name of each city. Our work is fundamentally different from theirs in that we use search query data to develop and test an aggregate measure of housing fear. Overall, Google search data can be used

---

[13]Da, Engelberg, and Gao (2012) build their index using search terms like "recession," "the great depression," and "credit card debt." They use their sentiment index to predict stock returns. Their results are similar to other studies that consider investor sentiment in the stock market.

for timely prediction or to capture previously unobserved behavior. Our Housing Distress Index extends the literature on both of these fronts in that it captures a new dimension of household behavior while predicting fundamental housing and mortgage credit indices.[14]

1   The Data

The search query terms that constitute the HDI are from Google Trends. Google is the most popular search engine in the United States. As of September 2011, Google accounted for 66 percent of all US internet searches.[15] Furthermore, according to the Pew Research Center, 92 percent of online adults use search engines.[16] This implies that queries through Google are representative of the US internet population.

To compile the search terms used in our index, we focus on phrases that contain a mortgage or housing related keyword in conjunction with a signal of distress. We begin by considering keywords, such as "mortgage" and "foreclosure," in combination with the word "help." The term "help" is the most commonly queried distress signal according to Google Adwords: over the 12 months prior to November 2012, U.S. internet users searched for the term "help" approximately 20.4 million times per month while queries for related terms such as "relief," "assistance," and "aid" totaled just 2.7 million, 5.0 million, and 9.1 million per month, respectively.[17]

Combining the word "help" with terms "mortgage" and "foreclosure" yields two obvious Google queries that can be immediately used as a starting point: "mortgage help" and "foreclosure help." Utilizing these two search terms as a basis for our index is advantageous as they are queried frequently by internet users: according to Google Adwords, during 2012, well after the housing crisis peaked, the phrases "foreclosure help" and "mortgage help" were queried 266,400 and 594,000 times respectively. This suggests that our housing distress index is the compilation of millions of internet search queries since

---

[14]The literature that employs Google search data is large, growing rapidly, and spans numerous fields. Recent applications include Baker and Fradkin (2011); Bollen, Mao, and Zeng (2011); Castle, Fawcett, and Hendry (2010); Goel et al. (2010), Mondria and Wu (2012), Schmidt and Vosen (2011), Stephens-Davidowitz (2011), and Carrière-Swallow and Labbé (2011).

[15]As measured by Hitwise, a company that monitors internet trends.

[16]See http://www.pewinternet.org/Reports/2011/Search-and-email/Report.aspx.

[17]Google Adwords reports the number of queries over a 12 month moving average for each search term. We do not consider the word "support" as this term is often queried in association with "child support" in a housing related context.

2004 and captures an extensively larger sample than the typical consumer survey.[18]

Entering "foreclosure help" and "mortgage help" into Google Trends produces a report that contains similar queries. We compile our list of search terms from those highlighted by Google on the condition that they contain a housing keyword and an indicator of distress. This process leads to the 16 search query terms listed in table 1.[19] All of the search terms contain a housing keyword along with the word "help" or "assistance." Thus, when an agent searches for one of the terms in table 1, he reveals a concern about his or others' housing-related financial distress: the agent may fear foreclosure, be behind on his mortgage payments, or be concerned about his ability to make future mortgage payments. For example, people may query "foreclosure help" or "government mortgage help" when they are searching for ways to stave off foreclosure or re-negotiate their mortgage payments.

In unreported results, we include a number of other search terms such as "foreclosure," "real estate foreclosures," and "foreclosures for sale." The results reported below are robust to the inclusion of those terms. Yet we omit those search terms that do not contain a signal of distress as people may query them, for example, to access information about foreclosures or if they are contemplating an investment in a foreclosed property.

For each search query term, Google Trends produces a search query index that ranges from 0 to 100.[20] The search query index increases as that term becomes more popular compared to all other search terms. Thus, each index from Google reaches a maximum value of 100 in the month that the popularity of the associated search term hits its peak. Zero values in the search query index represent months where so few searches

---

[18]The University of Michigan Survey Research Center, for example, surveys just 500 households per month to produce their Consumer Sentiment Index.

[19]To ensure that our results are not driven by any individual search term, we iteratively drop each search term, recompile the Housing Distress Index, and calculate the correlation coefficient with the original HDI. The minimum of these correlations is 0.998. We repeat this process and drop two search terms at a time. The minimum correlation coefficient in this case is 0.996. Together, these results imply that our index is robust to the exclusion of any individual search term.

[20]As noted by Choi and Varian (2012), the Google Trends data are computed from a sample of all internet searches. This implies that data from Google can "vary a few percent from day to day." As a robustness check, we re-downloaded the data for the HDI on different days. The minimum correlation coefficient between the HDIs constructed using data from different days was 0.98. The data used in this paper is available from the authors upon request.

occurred that a value for the index cannot be compiled.[21] In table 1, 12 of the 16 search terms contain more than one zero value. Yet the search terms that contain zero values also appear to capture information on housing distress.[22] For example, figures 2 and 3 display the screen snapshots from Google for the search terms "mortgage assistance" and "government mortgage help." Searches for the term "mortgage assistance" rise well before the start of the bear market in 2007M10 and peak just prior to the end of the bear market in 2009M03. On the other hand, searches for the term "government mortgage help" are almost nonexistent during the housing bubble, rise as the bear market begins, and then rocket up at the peak of the housing crisis. The other search query terms in table 1 that contain zero values also spike at the height of the housing crisis.

The right column in each panel of table 1 contains the date when each search query index reached its maximum value of 100. Of the 16 search terms, 13 reached their maximum value in 2009M02 or 2009M03; note that the bear market in equities also ended in 2009M03.[23] The search query index for the remaining three search terms in table 1 also spiked towards the end of the bear market. This implies that housing-related financial concerns reached their peak among all internet searches at the height of the housing crisis.

1.1 Housing Data

In the below analysis, we test for predictive effects of the HDI using quality-adjusted, repeat sales house price indices from both Federal Housing Finance Agency (FHFA) and Case-Shiller. The FHFA and Case-Shiller data differ in a number of ways. First, the FHFA dataset includes home purchase and refinance activity from 384 cities throughout the United States whereas the Case-Shiller data tracks just 20 cities. Indeed, many cities at the epicenter of the 2000s boom and bust (including numerous cities in California,

---

[21]More specifically, if there are too few searches for a given search term during a certain month, Google Trends automatically triggers a privacy filter and outputs a value of zero for the index. Zero values in our sample occur for certain search query terms during the housing bubble years.

[22]In a robustness check below, we exclude the 12 search terms that contain more than one missing value, re-compile our index, and calculate the correlation coefficient with the original HDI. This correlation coefficient is 0.986; suggesting that the search terms that contain missing values do not drive our results.

[23]The end of the bear market occurred on 2 March 2009 when the S&P500 hit its low point during recent crisis.

Arizona, and Florida) are not included in the Case-Shiller database. Those concerns are ameliorated in the much larger FHFA sample. However, the FHFA series is confined to sales or re-finance of houses using conventional, conforming mortgages, whereas the Case-Shiller series includes sales and re-finance of houses using all mortgage types, including subprime, Alt-A, jumbo, and the like. Furthermore, the FHFA only reports monthly house price time series for the national average and over the nine US Census divisions.[24] These Census divisions span multiple states and often cover large geographical areas. In comparison, the Case-Shiller data contain monthly information on 20 specific cities in addition to the national average. Hence, there is more limited cross-sectional variation among the FHFA divisions compared to the Case-Shiller cities. Moreover, the Case-Shiller house prices are calculated using a three month moving average; producing significant serial correlation. In fact, Ghysels et al. (2012) find that the AR(1) coefficient for the Case-Shiller national housing returns is 0.938; whereas that for the FHFA national housing returns is 0.756.[25] This makes prediction with regard to housing returns difficult after accounting for the serial correlation. In the end, each data set has its strengths and weaknesses. Yet as discussed below, our predictive results are similar whether we use the Case-Shiller or FHFA data.

The FHFA data are taken from the Federal Housing Finance Agency's website. We use the FHFA data that tracks house prices for US Census divisions and the nation overall. These data are seasonally adjusted and available at the monthly frequency. We transform the data into return form using the log first difference. In table 2, we report the summary statistics for the monthly FHFA housing returns including the mean, standard deviation, minimum, and maximum. In line with our expectations, the mean monthly return was lowest and the standard deviation was the highest for areas most affected by the housing boom and bust. For example, the Pacific division, which includes the state of California, produced the lowest mean monthly return of -0.165 percent and the highest standard deviation of 1.356 over our sample period. Less speculative areas, such as the

---

[24]Definitions of US Census divisions can be found on the Census Bureau's website: `https://www.census.gov/geo/www/us_regdiv.pdf`

[25]See also Case and Shiller (1989) for more about the serial correlation in house price returns.

West South Central division, which includes Texas, Oklahoma, Arkansas, and Louisiana, produced a positive mean monthly return of 0.210 percent with a standard deviation of just 0.659. The monthly returns on the national index, in the top row of table 2, were on average 0.00 percent over the sample period with a standard deviation of 0.673.

We download the Case-Shiller housing price data from the FRED economic database of the Federal Reserve Bank of St. Louis. The Case-Shiller data are all seasonally adjusted and available at the monthly periodicity. The Case-Shiller data are transformed into return series using the log first-difference. Table 3 displays the summary statistics for Case-Shiller housing returns for all 20 cities and for the National Index. The left most column identifies the city and the remaining columns in table 3 show the mean, the standard deviation, the minimum, and the maximum for returns across cities. The top row shows the summary statistics for the returns on the Case-Shiller National Housing Price Index. Over our sample, the mean monthly national return was -0.09 percent with a standard deviation of 1.078. The cities with the highest standard deviations were Phoenix, Las Vegas, Miami, Seattle, and Los Angeles, respectively. Hence, these cities represented the most speculative housing markets over the sample. As such, the monthly returns for Phoenix, the city with highest standard deviation, fluctuated markedly between -4.5 and 4.2 percent. Portland was the only city to produce positive mean monthly returns over our sample at 0.10 percent.

1.2 Other Data

We consider a number of other variables in our study. These include negative housing sentiment from the University of Michigan consumer survey; the VIX implied volatility index; the civilian employment-population ratio; the housing affordability index from the National Association of Realtors; the fraction of negative words in a popular Wall Street Journal column as in Tetlock (2007); retail sales; the yield curve; the Fed Funds rate; housing starts; foreclosure starts, prime and subprime delinquencies, and mortgage application volume from the Mortgage Bankers Association; and the Markit subprime ABX indices.

In their survey of consumer sentiment, the University of Michigan asks respondents

if "now is a good or bad time to buy a house and why." The survey further provides participants with an opportunity to explain their response. We focus on the reasons why "now is a bad time to buy a house." These include (1) prices are too high (*HighPrices*); (2) interest rates are too high (*HighRates*); (3) the respondent cannot afford a home (*CantAfford*); or (4) the future is uncertain (*UncertainFuture*). The respondent can choose more than one category. The data are monthly and available from the University of Michigan Survey Research Center website.

The VIX volatility index is from the Chicago Board Options Exchange (CBOE). The VIX index is the implied volatility of options on the S&P500 and is commonly referred to as the "investor fear gauge" for the stock market (Whaley (2000)).

The civilian employment series and the housing affordability index are available from the FRED economic database. The employment series is the civilian employment-population ratio and the housing affordability index measures the affordability of a mortgage for a median priced home given a median income buyer with a 20 percent down payment.

We also consider the fraction of negative words in the Wall Street Journal column "Abreast of the Market." Negative words are identified from "Abreast of the Market" using the Harvard IV-4 and Lasswell dictionaries and the General Inquirer software. A control for media pessimism is included in our study since negative news reports may affect agents' Google search behavior or capture other economic information. We construct the monthly pessimism series by averaging the daily pessimism series within each month.[26]

The retail sales, interest rate data, and housing starts information are from the FRED economic database. The yield curve measure utilized in this study is the 10-year Treasury Bond Rate minus the Fed Funds Rate. We also use the first difference in the Fed Funds Rate as an indicator for short term interest rates.

Foreclosure starts, prime and subprime delinquencies, and mortgage application volume are from the Mortgage Bankers Association. The foreclosure starts variable is the

---

[26]In June 2008, The Wall Street Journal stopped publishing "Abreast of the Market" daily. The column is now published weekly. Thus, after June 2008 we are averaging over a smaller number of observations.

number of loans that enter foreclosure in a given quarter, and prime and subprime delinquencies are the number of prime or subprime loans that become delinquent during a given quarter. The foreclosure and delinquency data are only available at the quarterly periodicity. Mortgage application volume represents the number of mortgage applications submitted for home purchases each month.

Lastly, we consider the ABX indices from Markit. The ABX indices track the market price of credit-default insurance on subprime mortgage-backed securities of different investment grades. For further details on the ABX indices, see Longstaff (2010).

2   The Housing Distress Index

We develop our Housing Distress Index (henceforth, HDI) from the internet search query terms listed in table 1. For each search query term, Google Trends produces a search query index that ranges from 0 to 100. As noted above, increases in the search query index indicate an increase in popularity relative to all other searches. To construct the HDI, we sum the search query indices and use the X12 algorithm from the US Census Bureau to seasonally adjust the series.[27] After removing the seasonality, we standardize the HDI to have zero mean and unit variance.

We first use all of the search query terms in table 1 to compile an "All Data HDI." Figure 4 displays the plot of the All Data HDI. The All Data HDI stays low while the housing bubble inflates, rises sharply through the subprime mortgage crisis, peaks in 2009M03, and then falls as the financial crisis abates.

While all search query terms are used to compute the HDI, 12 of the 16 terms in table 1 contain more than one zero value. These zero values represent months where so few searches occurred that the search query index could not be compiled. We drop search terms with more than one missing value and re-estimate our index to develop the "No Missing HDI."

Figure 5 displays the chart of the All Data HDI versus that of the No Missing HDI. The No Missing HDI stays low until 2007, rises as the housing bubble pops, peaks just

---

[27]As previously mentioned, some of the search terms, such as "government mortgage help," contain zero values during the bubble years. Summing the series is a parsimonious method for combining the Google Trends data that allows us to include all of the search terms listed in table 1.

prior to the end of the bear market, and then falls after 2009M03. The correlation between the two indices is 0.986. As such, the two indices are nearly identical. However, the All Data HDI contains more information as it is compiled from a larger number of search terms. Hence, in the remainder of the paper, we use the All Data HDI (henceforth, HDI).

In a standard Augmented Dickey-Fuller test we reject the null of a unit root in HDI at the 10 percent level with a p-value of 0.07. Yet to be conservative and avoid issues regarding persistence in regressions, as documented by Stambaugh (1999), we use the first difference in the HDI (henceforth, dHDI) in all of our predictive analyses. In unreported results, we test the predictive effects of the HDI in levels. Overall, these findings are consistent with our expectations and suggest that high levels in the HDI predict low future house price returns.

2.1 The orthogonalized dHDI

Although the HDI developed above closely corresponds with anecdotal accounts of housing pessimism, it may also be related to housing or macroeconomic fundamentals. To accommodate macroeconomic and housing indicators, we follow the stock sentiment literature (e.g. Baker and Wurgler (2006) and Lemmon and Portniaguina (2006)) and develop an orthogonalized housing distress index, $dHDI^{\perp}$. $dHDI^{\perp}$ is the set of residuals from a regression of dHDI on key housing variables including the log first-difference of the housing affordability index, housing starts, and mortgage application volume; and macroeconomic proxies including the log first-difference in industrial production, the civilian employment-population ratio, and retail sales; the first-difference in the Fed Funds rate; the yield curve; and media pessimism. We include the control for media pessimism as the tone of news reports may capture economic information and affect agents' Google search behavior.[28]

Overall, dHDI and $dHDI^{\perp}$ are closely related; suggesting that macroeconomic and housing fundamentals have a weak relationship with our housing distress measure. In-

---

[28]We do include the VIX in our set of controls as we use it as a dependent variable in later analyses. Yet when appropriate, we do include the VIX index as an additional control. By the Frisch-Waugh-Lovell theorem, the results below using the $dHDI^{\perp}$ are identical to using the original dHDI and including the macroeconomic and housing variables as additional controls.

deed, the R^2 from the aforementioned regression is just 0.15.[29] Thus, the macro and housing indicators explain little of the variation in the HDI. Since the macro and housing proxies have such low explanatory power, dHDI and dHDI$^\perp$ are highly correlated with a coefficient of 0.921.

In the analysis below, we report the results using both dHDI and dHDI$^\perp$. In general, the results are similar for both series. Yet our findings are slightly stronger when we use dHDI$^\perp$ in conjunction with the FHFA data, but slightly weaker for the Case-Shiller returns.[30] This suggests that orthogonalizing the HDI to various macro and housing indicators only adds noise to our regression models.

Together, the findings from this section imply dHDI and dHDI$^\perp$ are closely linked and that our housing distress index captures a new dimension of household behavior that is largely unrelated to macroeconomic proxies and housing market fundamentals.

3 The HDI and Consumer Sentiment

The above evidence, including the plot in figure 4 and the weak relationship between macro and housing fundamentals and dHDI, suggests that our index captures financial distress in the housing market. To further validate this claim, we compare the HDI to negative housing sentiment recorded in University of Michigan consumer sentiment survey responses. The monthly University of Michigan survey asks sampled households if now is a good or bad time to buy a house and why. As noted above, if the respondent answers that now is a bad time to buy a house, he can justify his choice citing (1) *HighPrices*, (2) *HighRates*, (3) *CantAfford*, or (4) *UncertainFuture*. Table 4 shows the correlations between these four categories and the HDI. As expected, the HDI is negatively correlated with *HighPrices*. Hence, the HDI is inversely related to consumers' concern about high house prices. Also, the HDI is positively correlated with *CantAfford*. This suggests that

---

[29]The F-statistic that tests the null hypothesis that all of the macro and housing variables are statistically different from zero is not significant at the ten percent level. Overall, the low explanatory power of the macro and housing variables is not surprising. The HDI rose slightly at the beginning of the sample while the other indicators showed no signs of economic distress. Then, during the crisis period, the spike in the HDI, which was over three standard deviations above the mean, was much larger than the variation in the other macro and housing variables.

[30]The results do diverge in a few cases when we use dHDI$^\perp$ as a substitute for dHDI. We specifically highlight these cases below.

the Housing Distress Index is also positively related to people's sentiment about their ability to afford a home. Most importantly, the correlation coefficient between the HDI and *UncertainFuture* is 0.740 and significant at the 1 percent level. Hence, the HDI is highly correlated with negative consumer sentiment about uncertain housing futures. Assuming as usual that agents dislike uncertainty, the high correlation between the HDI and *UncertainFuture* supports the claim that our measure captures respondent fear about the direction of the housing market.

Figure 6 shows the plot of the HDI versus *UncertainFuture*. The black vertical line in the figure represents the onset of the Lehman Crisis in September of 2008. *UncertainFuture* stays low at the start of the sample, rises sharply after the Lehman Crisis in September of 2008, and then fluctuates at elevated levels for the remainder of the sample. On the other hand, the HDI rises beginning in 2007, spikes in March 2009, and then declines as the crisis subsides. The plot in figure 6 suggests that (1) the HDI may be a leading indicator; (2) *UncertainFuture* reacted more to the Lehman Crisis; (3) the HDI soared at the height of the of the crisis while *UncertainFuture* did not; and (4) *UncertainFuture* remained elevated after the crisis even as the HDI fell. Together, these observations imply that the HDI and *UncertainFuture* are related, but capture different aspects of household behavior.

Next, table 5 shows the regressions of the first-difference of the sentiment components on dHDI_{t-1} and controls, while table 6 presents the results using the orthogonalized index, $\text{dHDI}_{t-1}^{\perp}$. In both tables, bootstrapped p-values are in parentheses. The controls include five lags of the dependent variable, one lag of the returns on the Case-Shiller 20 City National House Price Index; and one lag of the first-difference of the VIX index. Overall, the results are similar in both tables, highlighting the close link between our original index and its orthogonalized counterpart.

Increases in the HDI predict upward movement in all consumer explanations as to why now is a bad time to buy a house. Yet most noticeably, the predictive coefficients are large in magnitude and highly significant when *CantAfford* or *UncertainFuture* represent the dependent variable. These results match the plot in figure 6 and suggest that the

HDI acts as a leading indicator of negative housing market sentiment as it captures a more timely dimension of household distress.

4   Housing Distress and the ABX Indices

In this section, we compare the HDI to the ABX indices. As suggested by the Wall Street Journal, the ABX indices "are some of the most closely watched [subprime mortgage] barometers on Wall Street."[31] The ABX indices track the cost of credit-default swaps on subprime mortgage-backed securities of a certain investment grade.

The ABX indices fall as the cost to insure mortgage-backed securities rises. Hence, the ABX indices fall as investors become more bearish about housing and mortgage market performance. Note that the ABX indices reflect the mortgage and housing market pessimism of professional traders as only institutional investors participate in the trading of mortgage backed credit-default swaps. We use the ABX indices that began in 2006 (2006-01) as these data have the longest available time series.[32]

The advantage of comparing the HDI to subprime mortgage backed credit-default swaps via the ABX indices is that they are a tradable tool whose "...liquidity and standardization allows investors to accurately gauge market sentiment around the asset-class (subprime mortgage backed securities), and to take short or long positions accordingly."[33] Thus, if the HDI predicts the ABX indices (as we show below), it would suggest, via an efficient market hypothesis argument, that the HDI captures a dimension of household distress not previously observed by market participants.

We consider the AAA and AA ABX indices as the vast majority of subprime mortgage backed securities where rated either AAA or AA. According to Hull (2010), approximately 90 dollars of AAA rated securities were created from each 100 dollars of subprime mortgages. For example, in a specific case in 2006, Moody's rated three-quarters of $470 million in CDO securities as AAA. A full 80 percent of the CDOs' assets were subprime

---

[31]*The Wall Street Journal.* June 21, 2007. "Index With Odd Name Has Wall Street Glued; Morning ABX.HE Dose"

[32]In addition to the first half of 2006 (2006-01), the ABX indices were issued in the second half of 2006 (2006-02); the first half of 2007 (2007-01); and the second half of 2007 (2007-02). Over a common sample, the ABX indices are highly correlated for different investment grades.

[33]Markit ABX.HE product description. `http://www.markit.com/en/products/data/indices/structured-finance-indices/abx/abx.page`

mortgage-backed securities! This type of deal was typical for the time period.[34]

Figure 7 displays the plot of the AAA and AA ABX indices. The ABX indices are set to 100 on the day of issuance. The ABX indices start high at the beginning of the sample and then plummet in 2007 with the onset of the crisis. Not surprisingly, the AAA ABX performed significantly better than its lower-rated counterpart, falling by substantially less during the crisis period and then largely recovering in the aftermath.

Table 7 shows the correlations between the AAA and AA ABX returns and the dHDI. In both cases, the negative correlations with dHDI exceed 0.30 and are significant at the 1 percent level; suggesting that dHDI is closely related to the ABX returns.

Next, we build predictive regressions where the returns on the ABX indices represent the dependent variables. The model becomes

$$ABXret_{it} = \alpha_i + \beta_i \mathrm{dHDI}_{t-1} + \gamma_i Controls + u_{it} \tag{1}$$

where $ABXret_{it}$ is the return on the ABX index i, $i = \{$AAA, AA$\}$. The controls include five lags of the dependent variable and one lag of the following variables: the Case-Shiller National 20 City House Price returns, the first difference in the VIX index, and returns on the Dow Jones Industrial Average.

Table 8 shows the regression results. The coefficient on dHDI_{t-1} is negative and significant at the 10 percent level in both cases. Hence, increases in the HDI predict low future ABX returns. Moreover, the coefficient on dHDI is much larger and significant at the 1 percent level when the AA ABX returns are the dependent variable. This corresponds with our expectations and suggests that positive innovations in the HDI predict substantially lower returns for the AA ABX index with greater exposure to collateral loss. Moreover, the results are similar for the regressions that use the orthogonalized index, $\mathrm{HDI}^{\perp}$. This implies that the macroeconomic and housing proxies have little impact on the predictive relationship between the HDI and the ABX indices.

As the ABX indices reflect the prices on subprime mortgage credit default-swaps, the regression results in this section suggest that the HDI contains a key component of household distress not previously observed by market practitioners.

---

[34] "Triple-A Failure." *The New York Times*. April 27, 2008. `http://www.nytimes.com/2008/04/27/magazine/27Credit-t.html?pagewanted=all`

5 The HDI, Foreclosures, and Delinquencies

In this section, we evaluate the Housing Distress Index in relation to foreclosures and delinquencies. As noted above, the foreclosure and delinquency data are available only at quarterly lagged values from the Mortgage Bankers Association. In comparison, the HDI can be compiled at higher frequencies and can be updated in real time.

We first compare the HDI to foreclosure starts graphically. Foreclosure starts are the number of loans that enter foreclosure during a given quarter. Figure 8 displays the plot of the foreclosure starts (henceforth, *ForcStart*) versus the HDI. In figure 8, both series are standardized to have zero mean and unit variance. The black vertical line in the figure represents the start of Project Lifeline. Project Lifeline was a foreclosure prevention collaboration between the Treasury Department and the major mortgage lenders. Through this collaboration, mortgage banks worked with distressed borrowers to help them stay in their homes. The gray bar in the figure represents the California Foreclosure Prevention Act. The California Foreclosure Prevention Act placed a 90 day moratorium on foreclosures beginning on 1 June 2009. The dashed vertical line signifies the start of the foreclosure freeze following the announcement of the "robo-signing" scandal in October of 2010.[35]

As shown in figure 8, the HDI and foreclosure starts closely track one another during the first half of the sample. In the wake of Project Lifeline, however, foreclosure starts fell markedly. Hence, Project Lifeline appears to have reduced the incidence of lender foreclosure of defaulted borrowers. Yet as is evidenced from the chart, the HDI diverged from foreclosures and continued on its previous trend despite the onset of the federal program. The implication is that the federal program helped to stave off foreclosures, but did little to reduce housing distress. Similarly, foreclosure starts declined during the moratorium under the California Foreclosure Prevention Act in June 2009 and after the discovery of the "robo-signing" scandal in October 2010. The HDI, however, was already in decline when these events ensued and continued on its previous path.

In figure 8, we also plot the number of prime delinquencies and subprime delinquen-

---

[35] *The Wall Street Journal.* 9 October 2010. "U.S. Banks Get Boxed In on Foreclosures."

cies.[36] Clearly, the HDI acts as a leading indicator of these two measures as it rises faster during the housing collapse, peaks first during 2009Q1, and then falls as the crisis abates.

Next, we study the relationship between the HDI and foreclosures or delinquencies using predictive regressions. The delinquency data is broken down by category; prime and subprime. The dependent variables in these models are the growth in foreclosure starts or prime and subprime delinquencies. Table 9 displays the results. Here *HARP* is first difference in the number of loans issued under the Home Affordable Refinance Program (HARP) standardized to have zero mean and unit variance over the sample period; *lifeline* is a dummy variable representing the start of Project Lifeline; *caPreven* is a dummy variable representing the moratorium on foreclosures through the California Foreclosure Prevention Act; and *roboSign* is a dummy variable representing the foreclosure freeze following the announcement of the "robo-signing" scandal. Since the foreclosure and delinquency data are only available quarterly, the regressions in table 9 use only 29 observations. Thus, one must use caution when interpreting the results.

The left panel of the table holds the results using dHDI; the right panel shows our findings for the orthogonalized index, $dHDI^{\perp}$. Overall, the coefficients on dHDI are positive and significant. Hence, increases in the Housing Distress Index are associated with increases in future foreclosure starts, prime delinquencies, and subprime delinquencies. When we use the orthogonalized index, the coefficient on $dHDI^{\perp}$ is significant when the growth in foreclosure starts or prime delinquencies represents the dependent variable.

Overall, the results from this section indicate that the HDI leads traditional measures of housing market vulnerability; suggesting that households signal their financial concerns via Google search queries as captured by the HDI before entering into delinquency or foreclosure.

6   The HDI and the VIX

The VIX index often is described as an investor fear gauge for the stock market (Whaley (2000)). Figure 9 shows the plot of the VIX index versus the HDI. Both series are damped prior to the crisis and rise sharply in the context of market turmoil. However,

---

[36]The delinquency series are also standardized to have zero mean and unit variance.

while the VIX fluctuates before spiking during the Lehman Brothers crisis in September 2008, the HDI trends up markedly throughout the period of housing and mortgage market implosion. This suggests that the VIX reacts more to stock market events. The pattern between the two series is similar after the crisis, but the VIX index is more volatile.

The correlation coefficient between the VIX and the HDI is 0.85 and significant at the 1 percent level. Hence, the HDI is strongly related to the VIX. The HDI, however, is more highly correlated with negative housing market sentiment than is the VIX. The correlation between VIX and $UncertainFuture$ is 0.573. As noted above, the correlation between the HDI and $UncertainFuture$ is larger with a value of 0.740. The difference between the two correlation coefficients is statistically significant at the 1 percent level. Overall, these results suggest that distress in the housing market and expected risk in the stock market are related, but not identical.

Next, we describe the predictive relationship between these indices using the following regression model:

$$dVIX_t = \alpha + \beta \mathrm{dHDI}_{t-1} + \gamma Controls + u_t \tag{2}$$

The set of controls include three lags following variables: the first-difference of the VIX index, Case-Shiller 20 city national housing returns, returns on the Dow Jones Industrial Average, and a dummy variable for the Lehman bankruptcy. Table 10 shows the regression results. The model in column (2) uses the orthogonalized index, $\mathrm{dHDI}^{\perp}$.

In both cases, the coefficient on dHDI_{t-1} is positive and significant. Hence, increases in the HDI predict a rise in the VIX. Further, the coefficients on dHDI_{t-1} and $\mathrm{dHDI}^{\perp}_{t-1}$ are large in magnitude. Thus, a one standard deviation increase in the HDI, for example, leads to a 4.461 point increase in the VIX for the next month. This is a substantial and economically meaningful impact considering that the VIX reached its all time high in October 2008 at 89.53. Overall, the predictive power of the HDI is not surprising for this sample period as the housing crisis that began in 2007 led to extensive turmoil in the stock market in a contagious fashion (Longstaff (2010)).

7 Housing Distress and National Housing Returns

As noted above, figure 1 displays the plot of the Case-Shiller 20 City National Price Index versus the HDI. During the housing bubble period the Case-Shiller index rises sharply while the HDI remains relatively low. Once the crisis sets in, however, house prices plummet and the HDI shoots up. As the crisis abates, the HDI falls, but the Case-Shiller National Housing Price Index barely bounces off its lows. Thus, the relationship between housing distress and housing prices weakens in the aftermath of the crisis. From figure 1, it appears that the HDI and the Case-Shiller index are most closely linked during the crisis and that their relationship is asymmetric across crisis and non-crisis periods.

Table 11 shows the correlation coefficients between the dHDI or lagged dHDI and the FHFA or Case-Shiller national housing returns for the full sample (January 2004 - June 2011), for the non-crisis period (January 2004 - December 2006 and April 2009 - June 2011), and the crisis period (January 2007 - March 2009).[37] For the full sample, the correlation between dHDI and housing returns is relatively small but negative. Thus, increases in the HDI correlate with low housing returns. In the non-crisis period, the correlation coefficient is positive as the HDI was low but rising at the beginning of the sample while housing prices trended upwards. During the crisis, however, the HDI and housing returns are highly negatively correlated with a coefficient of -0.378 for the FHFA national housing returns -0.207 for the Case-Shiller returns. Moreover, the coefficients increase in magnitude during the crisis period when we employ the lagged dHDI: between January 2007 and March 2009, the correlation between $dHDI_{t-1}$ and the FHFA national returns is extremely large in magnitude at -0.598, while this coefficient for the Case-

---

[37]We use these dates for the crisis period throughout the remainder of the paper. Our results are not sensitive to these specific dates. In fact, many of our regression results become stronger when we use dates that more closely coincide with what many consider to be the start of the crisis. For example, one may argue that the housing crisis began in June 2007 when the Bear Sterns "High-Grade Structured Credit Strategies Enhanced Leverage Fund" first halted redemptions or in August of 2007 when BNP Paribus halted redemptions on three funds that invested in subprime securities. Using either of these dates as the start of the crisis only strengthens our findings reported below. Yet we show only our most conservative results and retain the January 2007 start date for the crisis to avoid data mining. For specific dates regarding the stoppage of redemptions for the Bear Sterns or BNP Paribus funds see the following articles from Businessweek and Bloomberg: `http://www.businessweek.com/stories/2007-06-12/bear-stearns-subprime-bathbusinessweek-business-news-stock-market-and-financial-advice` and `http://www.bloomberg.com/apps/news?pid=newsarchive&sid=aW1wj5i.vyOg`.

Shiller data also increases to -0.329. In general, these correlations suggest that dHDI and housing returns are related, especially during times of market turmoil, but that they are most closely linked when we consider the lagged version of our housing distress index.

Building on the correlations from table 11, we use the following regression model to formally describe the predictive relationship between dHDI and national housing returns:

$$\text{National}_t = \alpha + \beta_1 \text{dHDI}_{t-1} + \beta_2 \text{dHDI}_{t-1}^{crisis} + \gamma Controls + u_t \tag{3}$$

where $\text{dHDI}_{t-1}^{crisis} = \text{dHDI}_{t-1}$ during the crisis period and 0 otherwise. The crisis period is from January 2007 to March 2009. Table 12 shows the regression coefficients with bootstrapped p-values in parentheses. The controls include three lags of the national housing returns (the dependent variable) and three lags of the first difference of the VIX index. The left panel in table 12 shows the results for the FHFA data; the right panel uses Case-Shiller 20 City National Housing returns. In models (1) and (3), the functional form of the regression is identical to that described in equation 3; while models (2) and (4) use the orthogonalized index, $\text{dHDI}^{\perp}$. The bottom row in table 12 lists the sum of the coefficients on dHDI_{t-1} and $\text{dHDI}_{t-1}^{crisis}$, $\beta_1 + \beta_2$, with the p-value from the corresponding F-statistic in parentheses. This sum represents the total predictive effect of dHDI on returns during the crisis period.

When FHFA returns are the dependent variable, all of the coefficients are negative and significant at the 1 percent level. Hence, an increase in the HDI predicts low future returns for the entire sample period, but the effect is much larger in magnitude during the crisis period. More specifically, a one standard deviation increase in the HDI during the non-crisis period predicts a decrease in the FHFA national house prices of -0.28 percent per month. Yet in times of crisis, a one standard deviation increase in the HDI predicts a drop in prices of $-0.28 + -0.608 = -0.888$ percent per month. This result is large in magnitude, statistically significant at the 1 percent level, and economically meaningful as the minimum monthly return for the national FHFA house prices was -1.745 percent for our sample period. The predictive effects of dHDI during the crisis and non-crisis periods are also statistically different at the 1 percent level of significance.

23

The results in column (2) that use the orthogonalized index, $dHDI^{\perp}$, are similar but slightly stronger compared with those just discussed in column (1) as the total predictive effect of the HDI is slightly larger in the case of model (2) during the crisis. Overall, the comparison of models (1) and (2) suggests that accounting for macroeconomic proxies and housing market fundamentals has little effect on the predictive power of the HDI.

The right panel shows our findings using the Case-Shiller data. As previously mentioned, predicting the Case-Shiller housing returns is notably difficult as these series exhibit large and significant autocorrelation. Nevertheless, even after accounting for the serial correlation, the predictive effects of the HDI during the crisis period match our above findings using the FHFA data: the sum of the coefficients on $dHDI_{t-1}$ and $dHDI_{t-1}^{crisis}$ ($\beta_1 + \beta_2$) is large in magnitude at -0.852 and highly significant. Thus, during the crisis, an increase in the HDI predicts a drop in housing returns. In accordance with the findings above, these predictive results are economically meaningful as the minimum monthly Case-Shiller return over the sample was -2.069 percent.

In sum, the correlations in table 11 and the regression results in tables 12 highlight three important elements of the relationship between the HDI and housing returns: (1) $dHDI_{t-1}$ and housing returns are negatively correlated; (2) the HDI predicts national housing returns; and (3) the contemporaneous and predictive relationship between the HDI and housing returns is asymmetric and strongest during times of crisis. Overall, the asymmetric relationship between dHDI and housing returns is qualitatively similar to the relationship between the VIX index and stocks as the VIX is most negatively correlated with stock returns during times of market turmoil.[38] Hence, just as the VIX can be interpreted as a measure of fear in the stock market, so too might the HDI be interpreted as a fear gauge in the housing market.

---

[38]Szado (2009) finds that the VIX is most negatively correlated with large drops in the S&P500. This finding is consistent anecdotal accounts of market practitioners who use the VIX as a hedge for traditional investments during times of crisis. See the following article published by *Forbes*: `http://www.forbes.com/sites/kenrapoza/2011/08/07/in-crisis-mode-hedge-fund-managers-turn-to-the-vix/`

8 Housing Distress and the Cross-Section of Housing Returns

In this section, we examine the relationship between the HDI and returns over the cross-section for both the FHFA divisions and the Case-Shiller Cities in a panel regression. The panel data framework is advantageous as it allows us to pool the cross-sectional and time series data; this approach yields a large number of observations and increases the power of our statistical tests.

Within the panel data setting, our approach is similar to that described above as we consider the asymmetric predictive power of the HDI in the crisis and non-crisis periods. We use the same crisis dates across the panel such that the crisis period begins in January 2007 and ends in March 2009. Undoubtedly, there is heterogeneity over the cross-section as different regions fell victim to the crisis at difference times. Thus, not explicitly accounting for this heterogeneity may adversely affect our panel data findings. Indeed, in unreported results we specify a different crisis period for each city or Census division.[39] Incorporating these specifications into our panel data models produces larger predictive effects with regard to dHDI than those reported below. Yet we present our most conservative results here and maintain constant dates for the crisis period over the cross-section to avoid data mining. In later analyses, however, we do consider the heterogeneous predictive effects of the HDI over the cross-section using housing portfolios based on volatility and momentum; through this approach, we find that the predictive effects of our housing distress index are stronger for more speculative housing markets and for those in distress.

The panel regression model is as follows:

$$r_{it} = \gamma_i + \delta_m + \psi_y + \beta_1 \text{dHDI}_{t-1} + \beta_2 \text{dHDI}_{t-1}^{crisis} + \beta_3 Controls_{it} + v_{it} \tag{4}$$

where r_{it} is the return on city i at time t, γ_i represents division/city fixed effects, δ_m signifies month fixed effects, and ψ_y is the year fixed effects. The controls include lagged local house price returns.

---

[39]We define the crisis period for each division/city as a peak-to-trough episode in local house prices. This approach provides a practical method to capture the heterogeneity over the cross-section and follows the literature that attempts to date bear markets in the stock market. See, for example, Pagan and Sossounov (2003).

We follow Korniotis and Kumar (2012) and estimate equation 4 by OLS and calculate the p-values using the Driscoll and Kraay (1998) robust standard errors that accommodate both cross-sectional and serial correlation.

Table 13 shows the key regression output with the FHFA results in the left panel and the models using Case-Shiller data in the right panel. The bottom row in the table shows the total predictive effect of dHDI on returns during the crisis, $\beta_1 + \beta_2$.

For the FHFA data, all of the coefficients in table 13 are negative and significant at the 10 percent level. Moreover, the sum of the coefficients on dHDI_{t-1} and $\text{dHDI}_{t-1}^{crisis}$ ($\beta_1 + \beta_2$), which represents the total predictive effect during the crisis, is negative, highly significant, and large in magnitude. Thus, an increase in the HDI predicts a large drop in housing returns over the cross-section. Furthermore, the results are slightly stronger when we use the orthogonalized index, $\text{dHDI}^\perp$.

In the right panel of the table 13 that considers the Case-Shiller data, the coefficients that capture the predictive effects during the crisis period all have the expected sign, but are slightly smaller in magnitude.[40] When the HDI is used, the sum of the coefficients on $\text{dHDI}_{t-1} + \text{dHDI}_{t-1}^{crisis}$ is negative and significant at the ten percent level. Hence, for the cross-section of Case-Shiller cities, an increase in the HDI predicts a decrease in future returns.

8.1 Housing Distress and the Cross-Section of Housing Returns based on Volatility and Momentum

A disadvantage of the above panel data analysis is that the model estimates the average predictive effects over the cross-section. As already noted, this ignores the potential heterogeneous relationship between the HDI and returns for different cities or divisions. In this section, an alternative approach is used to compare the HDI to housing portfolios based on volatility or momentum. More specifically, we postulate that the predictive effects of the HDI are more pronounced for more speculative cities and for housing markets in distress. To test this hypothesis, we construct portfolios based on volatility and mo-

---

[40]This arises due to the large heterogeneity across Case-Shiller cities and the strong autocorrelation of returns especially at the local level. See Ghysels et al. (2012) and the references therein for further analysis on this latter point.

mentum using the FHFA and Case-Shiller data. For the volatility based portfolios with the Case-Shiller dataset, we form quartile portfolios based on the previous 2-24 month standard deviation of returns. As there are 20 Case-Shiller cities, five cities are placed in each bin. Likewise, the quartile momentum portfolios are based on returns over the previous 2-12 months where the top quartile comprises the housing markets with highest past returns. In a similar fashion, we build volatility and momentum portfolios with the data using the FHFA divisions. As there are just nine FHFA divisions, we place two divisions into the top and bottom bins for the volatility and momentum portfolios. Thus, 22.22 percent of the data are in the top and bottom volatility or momentum port-folios, respectively. Since the FHFA divisions cover such large geographical areas, there is relatively little variation across the sorted portfolios based on volatility or momentum.

The most volatile housing markets (those with the largest standard deviations of past returns) are the most speculative, while regions with the lowest momentum of past returns are in distress. Thus, these portfolios allow us to test for heterogeneous predictive effects of the HDI over the cross-section while accounting for specific housing market characteristics. This leads to the following regression model:

$$z_t = \alpha + \beta_1 \text{dHDI}_{t-1} + \beta_2 \text{dHDI}_{t-1}^{crisis} + \gamma Controls + v_t \tag{5}$$

where z_t is one of the portfolios based on housing volatility or momentum and the controls include three lags of the dependent variable and one lag of the first difference in the VIX index.

Table 14 lists the main regression output where the results in the left panel use FHFA data and findings in the right panel employ the Case-Shiller data. In the top rows of the table, the returns on the portfolios based on volatility (σ) are the dependent variable. More specifically, we estimate the model in equation 5 for the most volatile housing markets (the top quartile based on the 2-24 month standard deviation), the least volatile housing markets (the bottom quartile), and a long-short portfolio consisting of the returns on the most volatile markets less those that are least volatile. This long-short portfolio allows us to test for asymmetric predictive effects over the cross-section. For the Case-Shiller data, the total predictive effects of dHDI during the crisis are negative

27

and statistically significant across both the high and low volatility portfolios, but the sum of the regression coefficients on dHDI_{t-1} and $\text{dHDI}_{t-1}^{crisis}$ is nearly three times larger for the high volatility housing markets in the top row. This implies that increases in the HDI predict a larger drop in returns for the most speculative cities during times of market turmoil. To test for the asymmetric predictive effects over the cross-section, we let the High minus Low portfolio based on volatility be the dependent variable. In this case, the total effect during the crisis, the sum of β_1 and β_2, is also negative and statistically significant. Hence, the predictive effects of our housing distress index are larger in magnitude and more pronounced for more speculative cities. In general, the results are similar for FHFA data, but the sum of the coefficients on dHDI_{t-1} and $\text{dHDI}_{t-1}^{crisis}$ is not statistically significant when the High minus Low portfolio based on volatility is the dependent variable. This latter finding is not surprising as there is relatively less variation across the geographically large FHFA divisions.

In the bottom three rows of table 14, we report our findings for the momentum based portfolios. In line with our expectations, increases in the HDI predict a decrease in future returns for the housing markets with the lowest momentum, those in distress. These predictive effects are large in magnitude and statistically different from those for the highest momentum housing markets for the Case-Shiller data.[41] Hence, the relationship between the HDI and returns is asymmetric over the cross-section and stronger for housing markets in distress.

Table 15 shows the results using the orthogonalized index, $\text{dHDI}^{\perp}$. The signs of the regression coefficients are the same as those analyzed above, but magnitudes differ slightly. The magnitudes of the coefficients are larger for the FHFA data, but relatively smaller for the Case-Shiller data.

---

[41]The regression coefficients for the FHFA data are similar in sign and magnitude when the returns on the lowest momentum divisions are the dependent variable. Yet as there is relatively less variation across the FHFA divisions, when the High minus Low portfolio based on momentum is the dependent variable, the sum of the coefficients on dHDI_{t-1} and $\text{dHDI}_{t-1}^{crisis}$ is not significant. The sum of the coefficients, however, do retain the expected positive sign.

9 Interpretation of the Results: The HDI as a Housing Fear Gauge

Results thus far suggest the possible interpretation of our Housing Distress Index as a housing fear gauge. First, the HDI tracks anecdotal accounts of fear over the sample period as it remains low throughout the housing expansion but then spikes at the height of the crisis. As such, the HDI may be more a barometer of fear associated with implosion in housing markets than a measure of agent exuberance during a period of housing expansion. Second, the HDI is highly correlated with consumer sentiment regarding uncertain housing futures. This corresponds with our expectations as households divulge their concerns about current or future distress through the search query terms used to construct the HDI. Third, we find that the returns on the ABX indices and dHDI are highly inversely correlated. Recall that the ABX indices track the cost to investors to insure subprime mortgage-backed securities against default. As described above, the ABX indices plunged with the onset of the crisis as costs to insure mortgage-backed securities against default skyrocketed. The surge in cost of mortgage backed security related credit-default swaps was highly correlated to the spike in the HDI and indicative of fear of mortgage default among housing and mortgage investors. Hence, the high negative correlation between the HDI and the ABX indices suggests that the HDI is also related to the fear of mortgage default among institutional investors. Next, as discussed above, the HDI closely tracks the VIX stock market fear gauge. Since turmoil permeated through both the stock market and the housing market during the recent financial meltdown, the high correlation with the VIX further supports the interpretation of the HDI as a fear gauge. However, the VIX reacts more to stock market events while the HDI is more highly correlated with negative housing market sentiment. This implies that both the VIX and the HDI capture fear, but do so in different markets. Fifth, the relationship between the HDI and housing returns resembles that between the VIX and stock returns.

In addition, our index also appears to capture a new dimension of household behavior not previously documented in the literature. Not only does the HDI behave differently

over the sample period compared to other behavioral measures such as the survey based housing sentiment or the VIX index, but relatively little of its variation is explained by key housing market proxies or macroeconomic fundamentals. As such, dHDI is highly correlated with its orthogonalized counterpart, dHDI$^{\perp}$, and our key regression results only marginally change with the use of the orthogonalized measure.

Overall, the HDI appears to capture housing fear while at the same time overcoming some of the deficiencies in related indicators. Recall that (1) the University of Michigan consumer sentiment surveys do not ask sensitive mortgage related questions nor are they available in real time; (2) the ABX subprime mortgage CDS indices reflect the beliefs of institutional investors and do not directly measure the concerns of individual households; and (3) the VIX is more responsive to stock market events relative to those in the housing market. In comparison, the HDI utilizes broad-based, real time internet search query data that reflects agent negative sentiment and concern regarding housing and mortgage outcomes. Moreover, the HDI predicts the negative indicators considered in this study.

10   Conclusion

In this paper, we apply internet search query data to develop and test a broad-based, real-time gauge of distress in housing and mortgage markets. To do so, we first aggregate Google search queries for terms like "mortgage assistance" and "foreclosure help" to comprise a novel Housing Distress Index (HDI). We then assess the properties of the new index and its relationship to other housing and financial variables including survey based housing sentiment, foreclosures and delinquencies, the VIX index, and the ABX indices that track the cost of subprime mortgage credit-default swaps. Unlike these more common indicators, the HDI directly reflects negative agent sentiment revealed through search queries. This makes our index unique as the HDI captures a dimension of agent behavior not previously observed in the literature.

We apply the HDI to further examine the relationship between housing distress and various indicators of housing market performance. We find that the HDI predicts the ABX indices, foreclosures and delinquencies, national housing returns, and returns over the cross-section of cities. Results of the analysis reveal that increases in the HDI predict

a drop in national housing returns. Further, this effect is largest during times of housing market crisis. The asymmetric relationship between the HDI and national housing returns is qualitatively similar to that estimated between the VIX fear index and the stock returns as reported by Whaley (2000). Using both FHFA and Case-Shiller measures, we find that the HDI predicts returns over the cross-section, but the predictive effects are strongest for the most speculative housing markets and those in distress. Furthermore, we find that dHDI is inversely related to returns on ABX MBS credit-default swaps. Finally, increases in the HDI lead to future increases in foreclosures, delinquencies, and the VIX index. Research findings suggest the utility of internet search in the development of timely indices of housing sentiment. Further, particularly in times of crisis, indices such as the HDI may provide timely and important insights into housing and mortgage market performance.

A   Appendix: Tables

Table 1: Housing Distress search terms

Search Term	Max Date	Search Term	Max Date
foreclosure assistance	Mar-2009	home mortgage help	Mar-2009
foreclosure help government	Aug-2008	housing assistance	Jul-2009
foreclosure help	Feb-2009	mortgage assistance program	Mar-2009
government assistance mortgage	Mar-2009	mortgage assistance programs	Mar-2009
government mortgage help	Mar-2009	mortgage assistance	Mar-2009
help for mortgage	Mar-2009	mortgage foreclosure help	Nov-2008
help with mortgage	Mar-2009	mortgage government help	Mar-2009
home mortgage assistance	Mar-2009	mortgage help	Mar-2009

Notes: The search terms are compiled from Google Trends. In each panel, the column on the left holds the search term and the right column holds the date where the index for each search term reached its maximum.

Table 2: Summary Statistics for the FHFA housing return data

	Mean	Std	Min	Max
FHFA National	0.000	0.673	-1.791	1.040
East North Central	-0.116	0.689	-1.745	1.788
East South Central	0.101	0.853	-2.958	2.113
Middle Atlantic	0.147	0.741	-1.669	1.875
Mountain	-0.052	1.206	-3.412	2.122
New England	0.002	0.975	-2.910	2.704
Pacific	-0.165	1.356	-3.261	2.507
South Atlantic	-0.016	1.087	-2.786	2.825
West North Central	0.012	0.713	-2.140	1.532
West South Central	0.210	0.659	-1.814	1.520

Notes: This table provides summary statistics for FHFA housing returns data by Census division. The top row holds the national housing index across all divisions. All returns are in percentage form. The first column on the left shows the Census division; the remaining columns show the mean, standard deviation, minimum, and maximum of returns for each Census division.

Table 3: Summary statistics for the Case-Shiller housing return data

City	Mean	Std	Min	Max
National 20	-0.090	1.078	-2.069	1.607
Atlanta	-0.162	0.815	-2.393	1.215
Boston	-0.066	0.651	-1.635	1.393
Charlotte	-0.186	0.941	-4.119	2.684
Chicago	-0.220	0.998	-3.351	1.360
Cleveland	-0.004	0.627	-1.950	1.115
Dallas	-0.013	0.574	-1.448	1.624
Denver	-0.717	1.260	-4.289	1.569
Detroit	-0.034	0.496	-1.570	0.954
Las Vegas	-0.478	2.113	-4.645	5.322
Los Angeles	-0.054	1.573	-3.708	3.164
Miami	-0.185	1.821	-4.094	2.916
Minneapolis	-0.340	1.380	-5.198	3.107
New York	0.021	0.852	-1.807	1.714
Phoenix	-0.252	2.194	-4.497	4.221
Portland	0.100	1.073	-1.937	2.116
San Diego	-0.221	1.447	-3.419	3.263
San Francisco	0.096	1.051	-2.877	1.767
Seattle	-0.181	1.710	-4.428	2.882
Tampa	-0.182	1.553	-3.564	2.915
Washington DC	0.104	1.288	-2.316	2.638

Notes: This table provides summary statistics for Case-Shiller housing return data by city. The top row holds the national housing index across all 20 cities. All returns are in percentage form. The first column on the left shows the city. The remaining columns show the mean, standard deviation, minimum, and maximum of returns for each city.

Table 4: Correlation between the HDI and negative Michigan consumer housing sentiment

	HDI	HighPrices	HighRates	CantAfford	UncertainFuture
HDI	1.000				
HighPrices	-0.658	1.000			
HighRates	0.219	0.232	1.000		
CantAfford	0.685	-0.443	0.526	1.000	
UncertainFuture	0.740	-0.696	-0.049	0.495	1.000

Notes: Correlations between the HDI and the housing components of the University of Michigan Consumer Sentiment Survey. Respondents to the Michigan Sentiment survey are asked if now is a good or bad time to buy a home and why. In this table we consider the respondents' explanations of why now is a bad time to buy a home. The possible explanations include the following: (1) prices are too high (*HighPrices*); (2) interest rates are too high (*HighRates*); (3) the respondent cannot afford a home (*CantAfford*); or (4) the future is uncertain (*UncertainFuture*).

Table 5: Predictive regressions of Michigan negative housing sentiment on dHDI

	dHighPrices	dHighRates	dCantAfford	dUncertainFuture
(Intercept)	0.014	-0.002	0.01	0.000
	(0.747)	(0.955)	(0.869)	(1.000)
$dHDI_{t-1}$	0.135***	0.1***	0.267***	0.161***
	(0.000)	(0.01)	(0.000)	(0.028)
Controls	Included	Included	Included	Included
$RMSE$	0.923	0.959	0.871	0.746
R^2	0.257	0.2	0.342	0.506
AIC	232.662	239.053	223.039	197.302

Notes: Predictive regressions of the first difference of the components of the University of Michigan Consumer Sentiment Survey on the dHDI. Respondents to the Michigan Sentiment survey are asked if now is a good or bad time to buy a home and why. In this table we consider the respondents' explanations of why now is a bad time to buy a home. The possible explanations include the following: (1) prices are too high (*HighPrices*); (2) interest rates are too high (*HighRates*); (3) the respondent cannot afford a home (*CantAfford*); or (4) the future is uncertain (*UncertainFuture*). The controls include five lags of the dependent variable, one lag of the Case-Shiller 20 city national house price returns, and one lag of the first difference in the VIX index. Bootstrapped p-values are listed in parentheses. One, two, or three asterisks represent significance at the 15, 10, or 5 percent levels, respectively.

Table 6: Predictive regressions of Michigan negative housing sentiment on dHDI$^{\perp}$

	dHighPrices	dHighRates	dCantAfford	dUncertainFuture
(Intercept)	0.013	-0.003	0.009	-0.001
	(0.779)	(0.947)	(0.892)	(0.984)
$dHDI^{\perp}_{t-1}$	0.056***	0.08***	0.185***	0.107**
	(0.027)	(0.012)	(0.000)	(0.074)
Controls	Included	Included	Included	Included
$RMSE$	0.931	0.961	0.886	0.754
R^2	0.244	0.197	0.318	0.495
AIC	234.063	239.33	225.996	199.159

Notes: Predictive regressions of the first difference of the components of the University of Michigan Consumer Sentiment Survey on the dHDI. Respondents to the Michigan Sentiment survey are asked if now is a good or bad time to buy a home and why. In this table we consider the respondents' explanations of why now is a bad time to buy a home. The possible explanations include the following: (1) prices are too high (*HighPrices*); (2) interest rates are too high (*HighRates*); (3) the respondent cannot afford a home (*CantAfford*); or (4) the future is uncertain (*UncertainFuture*). The controls include five lags of the dependent variable, one lag of the Case-Shiller 20 city national house price returns, and one lag of the first difference in the VIX index. Bootstrapped p-values are listed in parentheses. One, two, or three asterisks represents significance at the 15, 10, or 5 percent levels, respectively.

Table 7: Correlations between dHDI and the returns on the ABX Indices

	dHDI	AAA ABX	AA ABX
dHDI	1.000		
AAA ABX	-0.411	1.000	
AA ABX	-0.307	0.797	1.000

Notes: Correlations between dHDI and the ABX index returns. dHDI is the first difference in the HDI. Each ABX index tracks the cost to insure a basket of 20 subprime mortgage backed securities. The ABX indices fall as investors become more pessimistic. The ABX indices are split up by investment rating.

Table 8: Predictive regressions of ABX returns on dHDI and controls

	(1) (AAA)	(2) (AAA)	(3) (AA)	(4) (AA)
(Intercept)	-0.118	-0.083	0.628	0.668
	(0.582)	(0.69)	(0.276)	(0.278)
$dHDI_{t-1}$	-2.100**	.	-7.03***	.
	(0.083)		(0.001)	
$dHDI_{t-1}^{\perp}$	.	-1.444*	.	-6.565***
		(0.143)		(0.004)
Controls	Included	Included	Included	Included
$RMSE$	2.755	2.778	8.729	8.768
R^2	0.544	0.536	0.469	0.465
AIC	302.929	303.958	441.337	441.861

Notes: Predictive regressions of the returns on the ABX indices on dHDI and controls. Each ABX index tracks the cost to insure a basket of 20 subprime mortgage backed securities. The ABX indices fall as investors become more pessimistic. The ABX indices are split up by investment rating. The regression equation for models (1) and (3) is $ABXret_{it} = \alpha + \beta_i dHDI_{t-1} + \gamma_i Controls + u_{it}$, where $ABXret_{it}$ is the return on ABX index i, $i = \{AAA, AA\}$. For the regressions in models (2) and (4), we use the first difference of the orthogonalized dHDI, $dHDI^{\perp}$. In each regression, we control for five lags of the dependent variable and one lag of the following variables: the Case-Shiller national 20 city housing returns, the first-difference in the VIX index, and the returns on the Dow Jones Industrial Average. Bootstrapped p-values are in parentheses. One, two, or three asterisks represent significance at the 15, 10, or 5 percent levels, respectively.

Table 9: Predictive regressions of foreclosures and delinquencies on dHDI and controls

	dHDI			dHDI$^{\perp}$		
	ForcStart	Prime Delin	Subprime Delin	ForcStart	Prime Delin	Subprime Delin
(Intercept)	6.287***	2.45***	3.653***	7.102***	2.896***	3.894***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
dHDI$_{t-1}$	9.468***	5.015***	2.711***	.	.	.
	(0.000)	(0.001)	(0.009)	.	.	.
dHDI$^{\perp}_{t-1}$	.	.	.	8.000***	1.443**	0.925
				(0.003)	(0.069)	(0.232)
HARP	5.657***	0.057	0.447	6.305***	-0.211	0.334
	(0.000)	(0.96)	(0.461)	(0.000)	(0.844)	(0.514)
lifeline	-5.271***	4.251***	0.351	-4.259**	4.616***	0.557
	(0.002)	(0.001)	(0.54)	(0.016)	(0.000)	(0.407)
caPreven	-0.323	0.774	-0.713	-2.409	-0.739	-1.509
	(0.935)	(0.738)	(0.677)	(0.514)	(0.676)	(0.309)
roboSign	-4.234***	-12.278***	-7.174***	-8.462***	-13.87***	-8.068***
	(0.043)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
$RMSE$	10.863	6.507	5.127	11.571	7.027	5.32
R^2	0.303	0.439	0.315	0.209	0.345	0.262
AIC	227.927	198.199	184.375	231.587	202.659	186.518

Notes: This table presents coefficients from predictive quarterly regressions on the log first difference of foreclosure starts (*ForcStart*), prime delinquencies (*Prime Delin*), or subprime delinquencies (*Subprime Delin*) on the dHDI or dHDI$^{\perp}$ and controls. *HARP* is the first difference in the number of loans issued under the HARP program standardized to have zero mean and unit variance over the sample period; *lifeline* is a dummy variable indicating the start of Project Lifeline; *caPreven* is a dummy variable that represents the foreclosure freeze initiated by the California Prevention Act; and *roboSign* is a dummy variable representing the announcement of the "robo-signing" scandal in October of 2010. Bootstrapped p-values are in parentheses. *RMSE* is the root mean-squared error. One, two, or three asterisks represent significance at the 15, 10, or 5 percent levels, respectively.

Table 10: Predictive regressions of the first difference of the VIX on dHDI and controls

	(1)	(2)
dHDI$_{t-1}$	4.461**	.
	(0.056)	
dHDI$^{\perp}_{t-1}$	.	3.882**
		(0.1)
Controls	Included	Included
$RMSE$	4.333	4.36
R^2	0.303	0.294
AIC	503.553	504.611

Notes: Predictive regressions of the first difference of the VIX index on dHDI or dHDI$^{\perp}$ and controls. dHDI$^{\perp}$ is the first difference of the HDI orthogonalized to various macroeconomic and housing factors. In each regression, we control for three lags of the following variables: the first difference of the VIX index, Case-Shiller 20 city national housing returns, the returns on the Dow Jones Industrial Average, and a dummy variable for the Lehman bankruptcy. Bootstrapped p-values are in parentheses. One, two, or three asterisks represents significance at the 15, 10, or 5 percent levels, respectively.

Table 11: Correlation coefficients between dHDI and the returns on the FHFA and Case-Shiller national housing price indices.

	Full Sample	Non-Crisis	Crisis
cor(dHDI,FHFAnat)	-0.084	0.3447	-0.378
cor(dHDI$_{t-1}$,FHFAnat)	-0.237	0.129	-0.598
cor(dHDI,CaseShiller)	-0.221	0.264	-0.207
cor(dHDI$_{t-1}$,CaseShiller)	-0.226	0.242	-0.329

Notes: This table shows the correlation coefficients between the first difference in the Housing Distress Index and the FHFA National Housing Price Returns or the Case-Shiller 20 city national housing returns for the full sample, and various sub-periods. We consider the full sample period from January 2004 to July 2011 (Full Sample); the non-crisis sample from 2004 to December 2006 and April 2009 to June 2011 (Non-Crisis); and the crisis period from 2007 to March 2009 (Crisis).

Table 12: Predictive regressions of FHFA or Case-Shiller national housing returns on dHDI and controls

	FHFA		Case-Shiller	
	(1)	(2)	(3)	(4)
(Intercept)	0.003	-0.018***	0.006	-0.021*
	(0.759)	(0.012)	(0.706)	(0.117)
dHDI_{t-1}	-0.28***	.	0.146	.
	(0.000)		(0.184)	
$\text{dHDI}_{t-1}^{crisis}$	-0.608***	.	-0.998***	.
	(0.002)		(0.001)	
$\text{dHDI}_{t-1}^{\perp}$	.	-0.368***	.	0.075
		(0.000)		(0.632)
$\text{dHDI}_{t-1}^{\perp,crisis}$	.	-0.548***	.	-0.864***
		(0.006)		(0.012)
Controls	Included	Included	Included	Included
$RMSE$	0.376	0.378	0.298	0.305
R^2	0.703	0.701	0.928	0.925
AIC	90.789	91.644	46.122	49.96
$\beta_1 + \beta_2$	-0.888***	-0.916***	-0.852***	-0.789***
	(0.010)	(0.026)	(0.004)	(0.020)

Notes: This table shows predictive regressions of the FHFA or the Case-Shiller national housing returns on dHDI and controls where dHDI is the first difference in the Housing Distress Index; dHDI^{crisis} equals dHDI during the crisis period and zero otherwise; and $\text{dHDI}^{\perp}$ is dHDI orthogonalized to various macroeconomic and housing factors. The regression equation in models (1) and (3) is as follows: $\text{National}_t = \alpha + \beta_1 \text{dHDI}_{t-1} + \beta_2 \text{dHDI}_{t-1}^{crisis} + \gamma Controls + u_t$, where National represents the returns on the FHFA or Case-Shiller national housing price index. Models (2) and (4) use $\text{dHDI}^{\perp}$. In all models, the controls include three lags of the dependent variable and three lags of the first difference in the VIX index. Bootstrapped p-values for the regression coefficients are listed in parentheses. The bottom row shows the total effect during a crisis, the sum of the coefficients, $\beta_1 + \beta_2$. The corresponding p-value from the F-statistic that tests the null hypothesis that the sum of the coefficients is equal to zero is listed in parentheses. One, two, or three asterisks represents significance at the 15, 10, or 5 percent levels, respectively.

Table 13: Predictive panel data regressions of FHFA divisions or Case-Shiller cities on dHDI and controls

	FHFA Data		Case-Shiller Data	
	(1)	(2)	(3)	(4)
dHDI_{t-1}	0.837***		-0.031	
	(0.000)		(0.763)	
$\text{dHDI}_{t-1}^{crisis}$	-1.500***		-0.177	
	(0.000)		(0.272)	
$\text{dHDI}_{t-1}^{\perp}$		0.796***		-0.012
		(0.000)		(0.909)
$\text{dHDI}_{t-1}^{\perp,crisis}$		-1.515**		-0.140
		(0.064)		(0.406)
Division/City Fixed Effects	Included	Included	Included	Included
Year Fixed Effects	Included	Included	Included	Included
Month Fixed Effects	Included	Included	Included	Included
Controls	Included	Included	Included	Included
Adj R^2	0.345	0.344	0.75	0.755
$\beta_1 + \beta_2$	-0.663***	-0.719***	-0.208**	-0.152
	(0.000)	(0.008)	(0.058)	(0.223)

Notes: This table presents the predictive panel data regressions of returns on the FHFA/Census divisions or Case-Shiller cities on dHDI and controls. dHDI is the first difference in the HDI and $\text{dHDI}^{\perp}$ is the first difference in the HDI orthogonalized to various macroeconomic factors. dHDI^{crisis} equals dHDI during the crisis period and zero otherwise. The regression equation for models (1) and (3) is $r_{it} = \gamma_i + \delta_m + \psi_y + \beta_1 \text{dHDI}_{t-1} + \beta_2 \text{dHDI}_{t-1}^{crisis} + \beta_3 Controls_{it} + v_{it}$. In models (2) and (4), we use $\text{dHDI}_{t-1}^{\perp}$. γ_i, δ_m, and ψ_y represent Division/City fixed effects, month fixed effects, and year fixed effects, respectively. Controls include lagged returns for each city. The p-values listed in parentheses are based on the Driscoll and Kraay (1998) robust standard errors that accommodate both serial and cross-sectional correlation. The bottom row shows the total effect during a crisis, the sum of the coefficients, $\beta_1 + \beta_2$. The corresponding p-value from the F-statistic that tests the null hypothesis that the sum of the coefficients is equal to zero is listed in parentheses. One, two, or three asterisks represents significance at the 15, 10, or 5 percent levels, respectively.

Table 14: Predictive Regressions of volatility or momentum portfolios based on FHFA divisions on Case-Shiller cities on dHDI and controls

		FHFA			Case-Shiller		
		dHDI_{t-1}	$\text{dHDI}^{crisis}_{t-1}$	$\beta_1 + \beta_2$	dHDI_{t-1}	$\text{dHDI}^{crisis}_{t-1}$	$\beta_1 + \beta_2$
σ	High	0.103	-1.137***	-1.034**	0.437**	-1.402***	-0.966***
		(0.357)	(0.001)	(0.077)	(0.077)	(0.013)	(0.035)
σ	Low	-0.031	-0.636***	-0.666*	-0.109***	-0.215*	-0.324*
		(0.840)	(0.021)	(0.109)	(0.036)	(0.110)	(0.128)
σ	High − Low	0.159	-0.728***	-0.569	0.445***	-1.149***	-0.704**
		(0.311)	(0.025)	(0.321)	(0.032)	(0.000)	(0.092)
Mom	High	-0.108	-1.052***	-1.16***	-0.077	-0.323**	-0.401
		(0.474)	(0.002)	(0.019)	(0.488)	(0.066)	(0.185)
Mom	Low	-0.104	-1.058***	-1.163***	0.011	-1.203**	-1.192***
		(0.189)	(0.000)	(0.021)	(0.964)	(0.058)	(0.001)
Mom	High − Low	0.037	0.141	0.178	-0.065	0.826**	0.761*
		(0.683)	(0.486)	(0.737)	(0.814)	(0.018)	(0.118)

Notes: This table shows the predictive regressions of various volatility or momentum portfolios on dHDI and controls. The results using the FHFA data are in the left panel; the Case-Shiller findings are in the right panel. dHDI is the first difference in the HDI and dHDI^{crisis} equals dHDI during the crisis period and zero otherwise. The regression model is $z_t = \alpha + \beta_1 \text{dHDI}_{t-1} + \beta_2 \text{dHDI}^{crisis}_{t-1} + \gamma Controls + \upsilon_t$ where z_t is any one of the portfolios based on volatility (σ) or momentum (Mom). For the Case-Shiller data, the housing returns from a given city have high (low) volatility if their previous 2-24 month standard deviation of returns is in the upper (lower) 25 percent, while a city has high (low) momentum if its returns over the previous 2-12 months are in the upper (lower) 25 percent. A similar definition applies for the FHFA data, but an FHFA division has high (low) volatility or momentum if it is in the upper (lower) 22.22 percent. The controls include three lags of the dependent variable and one lag of the first difference of the VIX index. Bootstrapped p-values for the regression coefficients are listed in parentheses. The right column in each panel shows the total effect during a crisis, the sum of the coefficients, $\beta_1 + \beta_2$. The corresponding p-value from the F-statistic that tests the null hypothesis that sum of the coefficients is equal to zero is listed in parentheses. One, two, or three asterisks represents significance at the 15, 10, or 5 percent levels, respectively.

Table 15: Predictive Regressions of volatility or momentum portfolios based on FHFA divisions on Case-Shiller cities on dHDI$^\perp$ and controls

		FHFA			Case-Shiller		
		dHDI$^\perp_{t-1}$	dHDI$^{\perp,crisis}_{t-1}$	$\beta_1 + \beta_2$	dHDI$^\perp_{t-1}$	dHDI$^{\perp,crisis}_{t-1}$	$\beta_1 + \beta_2$
σ	High	0.148	-1.578***	-1.431***	0.355	-1.001**	-0.646
		(0.212)	(0.000)	(0.039)	(0.185)	(0.078)	(0.220)
σ	Low	-0.018	-0.709***	-0.726*	-0.164***	-0.029	-0.193
		(0.885)	(0.003)	(0.139)	(0.007)	(0.820)	(0.438)
σ	High − Low	0.179*	-0.837***	-0.658	0.441***	-0.898***	-0.457
		(0.139)	(0.030)	(0.336)	(0.036)	(0.007)	(0.338)
Mom	High	-0.272***	-1.073***	-1.346***	-0.042	-0.206	-0.248
		(0.029)	(0.002)	(0.023)	(0.681)	(0.177)	(0.493)
Mom	Low	-0.147**	-0.995***	-1.142**	-0.09	-0.864	-0.953***
		(0.061)	(0.000)	(0.052)	(0.746)	(0.164)	(0.025)
Mom	High − Low	-0.075	0.129	0.054	0.038	0.704**	0.743
		(0.530)	(0.669)	(0.932)	(0.883)	(0.080)	(0.199)

Notes: This table shows the predictive regressions of various volatility or momentum portfolios on dHDI$^\perp$ and controls. The results using the FHFA data are in the left panel; the Case-Shiller findings are in the right panel. dHDI$^\perp$ is the orthogonalized dHDI and dHDI$^{\perp,crisis}$ equals dHDI$^\perp$ during the crisis period and zero otherwise. The regression model is $z_t = \alpha + \beta_1 \text{dHDI}^\perp_{t-1} + \beta_2 \text{dHDI}^{\perp,crisis}_{t-1} + \gamma Controls + v_t$ where z_t is any one of the portfolios based on volatility (σ) or momentum (Mom). For the Case-Shiller data, the housing returns from a given city have high (low) volatility if their previous 2-24 month standard deviation of returns is in the upper (lower) 25 percent, while a city has high (low) momentum if its returns over the previous 2-12 months are in the upper (lower) 25 percent. A similar definition applies for the FHFA data, but an FHFA division has high (low) volatility or momentum if it is in the upper (lower) 22.22 percent. The controls include three lags of the dependent variable and one lag of the first difference of the VIX index. Bootstrapped p-values for the regression coefficients are listed in parentheses. The right column in each panel shows the total effect during the crisis, the sum of the coefficients, $\beta_1 + \beta_2$. The corresponding p-value from the F-statistic that tests the null hypothesis that sum of the coefficients is equal to zero is listed in parentheses. One, two, or three asterisks represents significance at the 15, 10, or 5 percent levels, respectively.

B Appendix: Figures

Figure 1: The HDI vs the Case-Shiller index
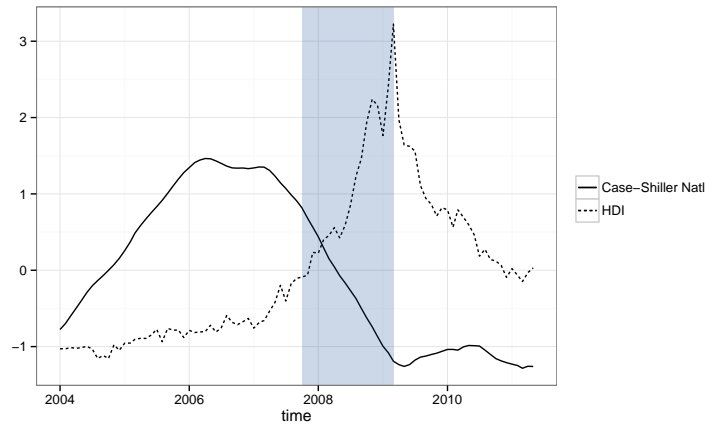
Notes: The Housing Distress Index (HDI) versus the Case-Shiller 20 City National Housing Price Index. The Case-Shiller 20 City National Housing Price Index is the solid line; the Housing Distress Index is the dotted line. Both series are seasonally adjusted and are transformed to have mean zero and unit variance. The blue shaded area indicates a bear market defined as a 20 percent or more drop in the S&P500 over a period of two or more months.

Figure 2: Searches for "mortgage assistance"
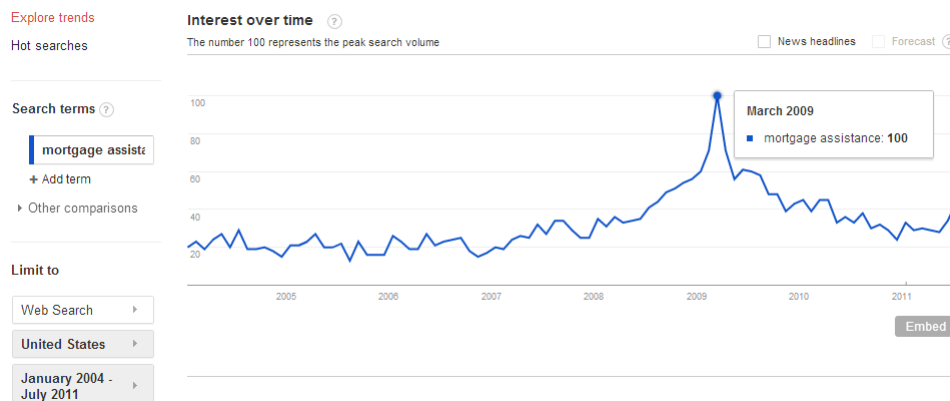
Notes: Google searches for "mortgage assistance" from Google Trends.

Figure 3: Searches for "government mortgage help"
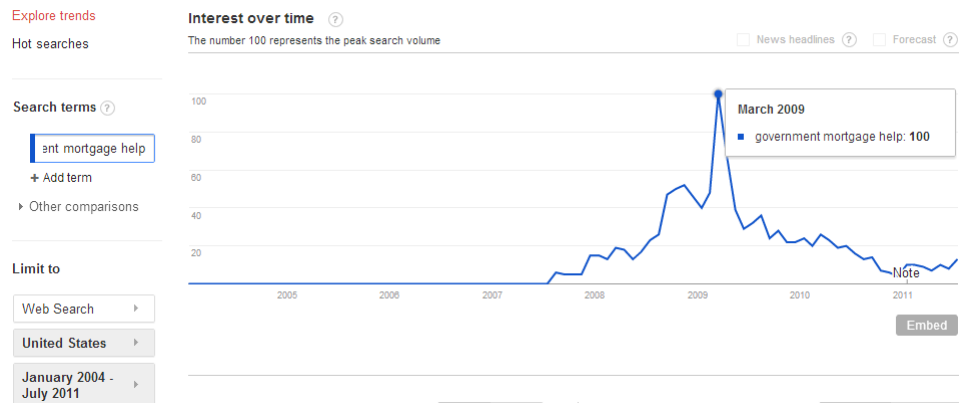
Notes: Google searches for "government mortgage help" from Google Trends.

Figure 4: The All Data HDI
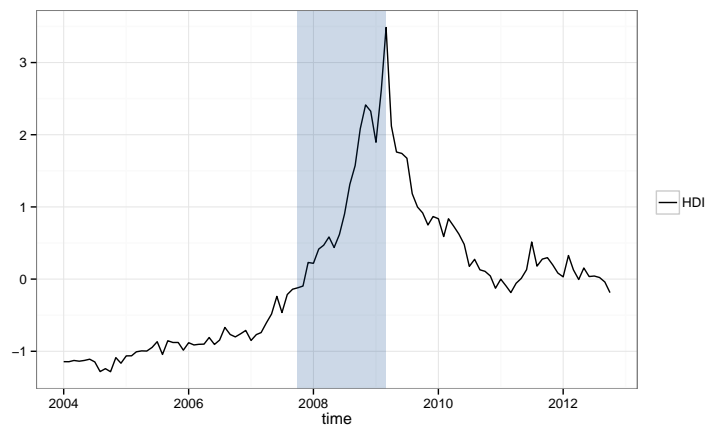
Notes: The All Data Housing Distress Index. The All Data Housing Distress index uses all search query terms. The blue shaded area indicates a bear market defined as a 20 percent or more drop in the S&P500 over a period of two or more months.

43

Figure 5: The All Data HDI versus the No Missing HDI
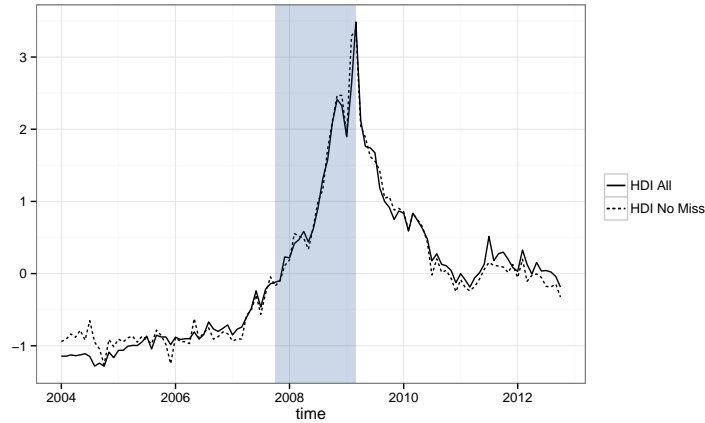
Notes: The All Data Housing Distress Index (solid line) and the No Missing Housing Distress Index (dotted line). The All Data Housing Distress index uses all search query terms; the No Missing Housing Distress Index uses search terms with data available over the entire sample period. The blue shaded area indicates a bear market defined as a 20 percent or more drop in the S&P500 over a period of two or more months.

Figure 6: The HDI versus *UncertainFuture*
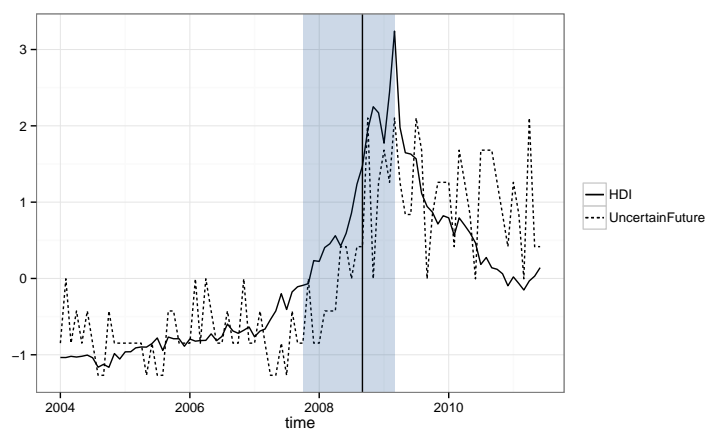
Notes: The HDI (solid line) versus *UncertainFuture* (dotted line). The blue shaded area indicates a bear market defined as a 20 percent or more drop in the S&P500 over a period of two or more months. The black vertical line marks the Lehman bankruptcy in September 2008.

Figure 7: The ABX 2006-01 prices
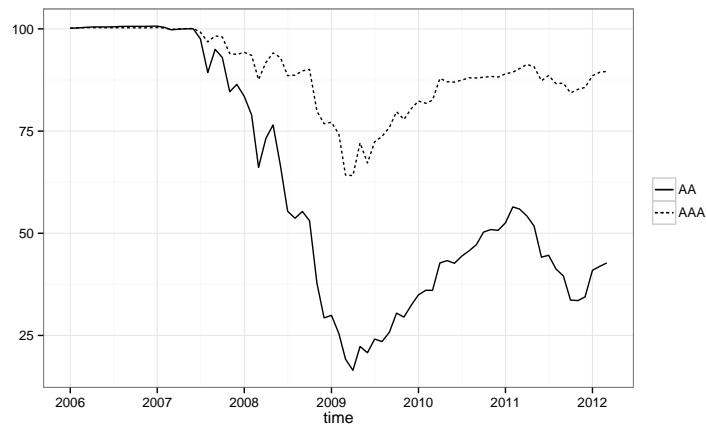
Notes: Plots of the ABX indices issued in the first half of 2006 (2006-01). The ABX indices are set to a value of 100 on the day of issuance. Each ABX index tracks the cost to insure a basket of 20 subprime mortgage backed securities. The ABX indices fall as investors become more pessimistic. The ABX indices are split up by investment rating.

Figure 8: The HDI versus Foreclosure Starts and Delinquencies
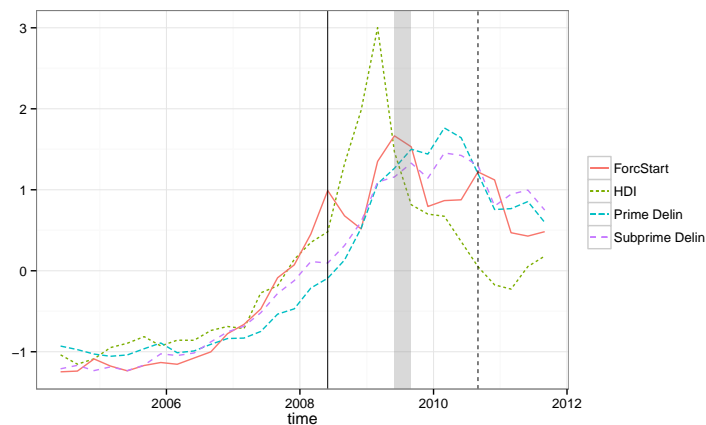
Notes: The Housing Distress Index (green, dotted line) versus foreclosure starts (*ForcStart*; red, solid line), prime delinquencies (*Prime Delin*; blue, dashed line), and subprime delinquencies (*Subprime Delin*; purple, dashed line). Foreclosure starts are the number of loans that enter foreclosure during a given quarter; prime delinquencies are the number of prime loans that become delinquent during a given quarter; and subprime delinquencies are the number of subprime loans that become delinquent during a given quarter. All series are transformed to have mean zero and unit variance. The data are quarterly. The solid vertical line marks the start of "Project Lifeline." The gray shaded area marks the foreclosure moratorium under the California Foreclosure Prevention Act. The dashed black vertical line represents the announcement of the "robo-signing" scandal in the October of 2010.

Figure 9: The HDI versus the VIX Index
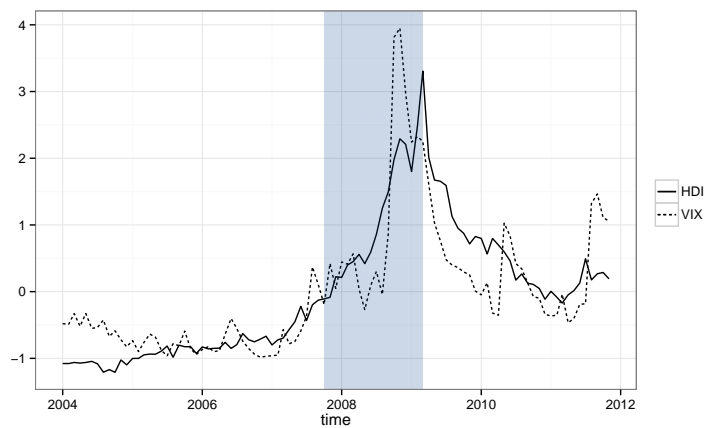
Notes: The HDI versus the VIX index. The solid line represents the Housing Distress Index; the dashed line represents the VIX index. Both series are monthly. We transform the VIX index to monthly data by taking the average within each month. The blue shaded area indicates a bear market defined as a 20 percent or more drop in the S&P500 over a period of two or more months.

References

[1] A. Ang, R. J. Hodrick, Y. Xing, and X. Zhang. The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1):259–299, 2006.

[2] M. Baker and J. Wurgler. Investor sentiment and the cross-section of stock returns. *Journal of Finance*, 61(4):1645–1680, 2006.

[3] S. Baker and A. Fradkin. What drives job search? Evidence from Google search data. *Working Paper*, 2011.

[4] E. Beracha and M. B. Wintoki. Forecasting residential real estate price changes from online search activity. *Journal of Real Estate Research*, Forthcoming.

[5] J. Bollen, H. Mao, and X.-J. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, Mar. 2011.

[6] J. Y. Campbell and L. Hentschel. No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, 31(3):281–318, 1992.

[7] Y. Carrière-Swallow and F. Labbé. Nowcasting with Google Trends in an emerging market. *Journal of Forecasting*, 2011. Working Papers Central Bank of Chile 588.

[8] K. E. Case and R. J. Shiller. The efficiency of the market for single-family homes. *American Economic Review*, 79(1):125–37, 1989.

[9] J. L. Castle, N. W. P. Fawcett, and D. F. Hendry. Nowcasting is not just contemporaneous forecasting. *National Institute Economic Review*, 201(1):71–89, Oct. 2009.

[10] H. Choi and H. Varian. Predicting the Present with Google Trends. *Economic Record*, 88(s1):2–9, 2012.

[11] G. Conti and E. Sobiesk. An honest man has nothing to fear: User perceptions on web-based information disclosure. In *SOUPS*, volume 7, pages 112–121, 2007.

[12] Z. Da, J. Engelberg, and P. Gao. In search of attention. *The Journal of Finance*, 66(5):1461–1499, 2011.

[13] Z. Da, J. Engelberg, and P. Gao. The sum of all fears: Investor sentiment and asset prices. *Working Paper*, 2012.

[14] J. C. Driscoll and A. C. Kraay. Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics*, 80(4):549–560, 1998.

[15] K. R. French, G. W. Schwert, and R. F. Stambaugh. Expected stock returns and volatility. *Journal of financial Economics*, 19(1):3–29, 1987.

[16] K. D. Garbade. Federal reserve margin requirements: A regulatory initiative to inhibit speculative bubbles. *Crises in the Economic and Financial Structure, Lexington, MA: Lexington Books*, 1982.

[17] E. Ghysels, A. Plazzi, W. Torous, and R. Valkanov. Forecasting real estate prices. *Handbook of Economic Forecasting*, Forthcoming.

[18] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014, 2009.

[19] S. Goel, J. M. Hofman, S. Lahaie, D. M. Pennock, and D. J. Watts. Predicting consumer behavior with web search. *Proceedings of the National Academy of Sciences*, 7(41):17486–17490, Sep 27 2010.

[20] J. Hull. Credit ratings and the securitization of subprime mortgages. *Working Paper*, 2010.

[21] N. Kiyotaki and J. Moore. Credit cycles. *The Journal of Political Economy*, 105(2):211–248, 1997.

[22] G. Korniotis and A. Kumar. State-level business cycles and local return predictability. *Journal of Finance, Forthcoming*, 2012.

[23] O. Lamont and J. Stein. Leverage and house-price dynamics in us cities. *RAND Journal of Economics*, 21:498–514, 1999.

[24] M. Lemmon and E. Portniaguina. Consumer confidence and asset prices: Some empirical evidence. *Review of Financial Studies*, 19(4):1499, 2006.

[25] F. A. Longstaff. The subprime credit crisis and contagion in financial markets. *Journal of Financial Economics*, 97(3):436–450, 2010.

[26] J. Mondria and T. Wu. Familiarity and surprises in international financial markets: Bad news travels like wildfire, good news travels slow. *Working Paper*, 2011.

[27] J. Mondria, T. Wu, and Y. Zhang. The determinants of international investment and attention allocation: Using internet search query data. *Journal of International Economics*, 82(1):85–95, 2010.

[28] A. R. Pagan and K. A. Sossounov. A simple framework for analysing bull and bear markets. *Journal of Applied Econometrics*, 18(1):23–46, 2003.

[29] T. Schmidt and S. Vosen. Forecasting private consumption: Survey-based indicators vs. Google Trends. *Journal of Forecasting*, 20(6):565–578, Sept. 2011.

[30] A. Shleifer and R. W. Vishny. Liquidation values and debt capacity: A market equilibrium approach. *The Journal of Finance*, 47(4):1343–1366, 1992.

[31] E. Singer. The use of incentives to reduce nonresponse in household surveys. *Survey Nonresponse*, pages 163–177, 2002.

[32] R. F. Stambaugh. Predictive regressions. *Journal of Financial Economics*, 54(3):375–421, 1999.

[33] J. C. Stein. Prices and trading volume in the housing market: A model with down-payment effects. *The Quarterly Journal of Economics*, 110(2):379–406, 1995.

[34] S. Stephens-Davidowitz. The effects of racial animus on voting: Evidence using Google search data. *Working Paper*, 2011.

[35] E. Szado. Vix futures and options-a case study of portfolio diversification during the 2008 financial crisis. *Working Paper*, 2009.

[36] P. C. Tetlock. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168, 2007.

[37] R. E. Whaley. The investor fear gauge. *The Journal of Portfolio Management*, 26(3):12–17, 2000.