Journal of Urban Economics 96 (2016) 91–111

Contents lists available at ScienceDirect

Journal of Urban Economics

journal homepage: www.elsevier.com/locate/jue

Mortgage default risk: New evidence from internet search queries[☆]

Marcelle Chauvet[a], Stuart Gabriel[b,*], Chandler Lutz[c]

[a] University of California, Riverside, United States
[b] University of California, Los Angeles, United States
[c] Copenhagen Business School, Denmark

ARTICLE INFO

Article history:
Received 7 October 2014
Revised 29 July 2016
Available online 31 August 2016

JEL Classification:
R30
G02
G12

Keywords:
Mortgage default risk

ABSTRACT

We use Google search query data to develop a broad-based and real-time index of mortgage default risk. Unlike established indicators, our Mortgage Default Risk Index (MDRI) directly reflects households'concerns regarding their risk of mortgage default. The MDRI predicts housing returns, mortgage delinquency indicators, and subprime credit default swaps. These results persist both in- and out-of-sample and at multiple data frequencies. Together, research findings suggest internet search queries yield valuable new insights into household mortgage default risk.

1. Introduction

The financial concerns of homeowners are of paramount importance to the US economy.[1] As evidenced during the 2000s housing and financial crisis, elevated mortgage delinquencies and defaults can dampen future house prices, exacerbate episodes of pessimism among consumers and investors, and wreak havoc on the macroeconomy and financial markets. While aggregate financial risk was captured by an array of generalized market indices such as the VIX index (the so-called "stock market fear index" as proxied by the expected variance of S&P500 stock returns), none of these measures provided timely insights specific to mortgage default risk during run-up to the crisis. Further, the few available measures of mortgage default risk only captured information known to lenders or financial market participants and thus neglected potentially sensitive information on mortgage distress emanating directly from households. The paucity of information on household mortgage concerns was striking, given the leading role of housing in the global downturn.

In this paper, we aim to fill this gap via the development and predictive test of a new and direct measure of mortgage default risk. We seek a broad-based, real-time gauge capable of capturing the marked swings in default risk that pervaded housing markets over recent years. In developing and testing such an index, we apply new, internet search query information from Google. Google search queries can be tabulated in "real-time" (Choi and Varian, 2012) at multiple frequencies. In contrast, popular housing indices from Case-Shiller, for example, are only available with a 2 month lag, while other potential drivers of mortgage default, such as unemployment rates, have little predictive power (Gyourko and Tracy, 2014).

Using Google data, we collect sensitive information directly from individuals seeking assistance via internet search on issues of mortgage default and home foreclosure.[2] Specifically, we aggregate Google search queries for terms like "foreclosure help" and "government mortgage help" to compile a novel Mortgage Default Risk Index (henceforth, MDRI). While these and related searches are derived from all households, a universe that includes both owners and renters, the bulk of such searches likely emanate from property owners as they are most likely to be concerned with mortgage default.[3] We infer that when a user seeks help via a Google search

☆ We would like to thank Brian Bucks, Ed Glaeser, Aurel Hizmo, Ed Leamer, Selale Tuzel, Cindy Soo, Todd Sinai, Seth Stephans-Davidowitz, Hal Varian, two anonymous referees, an editor, and conference participants at the American Economic Association, the NBER, the Summer Real Estate Symposium (at the WFA), the UCLA/Federal Reserve Bank of San Francisco Conference on Housing and the Macroeconomy, the UCLA Finance Department, the AREUEA International Conference, IDC Herzeliya, Tel Aviv University, and Copenhagen Business School for their helpful comments.

* Corresponding author.
E-mail addresses: chauvet@ucr.edu (M. Chauvet), stuart.gabriel@anderson.ucla.edu (S. Gabriel), cl.eco@cbs.dk (C. Lutz).

[1] There are over 45 million home mortgages outstanding in the United States valued at nearly 10 trillion dollars. 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.

[3] We would like to thank the editor for this point.

http://dx.doi.org/10.1016/j.jue.2016.08.004
0094-1190/

query such as "mortgage foreclosure help," she *divulges* her concern about mortgage failure or foreclosure. This makes the MDRI unique compared to other proxies for mortgage default risk – we simply let households speak and then capture the pertinent information related to their concerns about mortgage default. Other housing information, such as housing sentiment surveys by organizations such as the University of Michigan, only ask households if now is a good or bad time to buy a home, but do not ask sensitive questions about mortgage default.[4] Moreover, proxies such as mortgage delinquencies or foreclosures only reflect information tabulated from lenders and thus may miss important incremental information pertaining to "at risk" households. Further, these indices are also only available at substantial lags.

Our index is related to the theory and empirics of mortgage default. According to the "double trigger hypothesis", only a substantial loss of income, via unemployment for example, will assure mortgage default when a borrower has negative equity (Foote et al., 2008).[5] A borrower who has positive equity and experiences an adverse income shock can simply sell the house, pay off the mortgage, and reap the proceeds. Mortgage datasets, however, typically do not observe contemporaneous job loss or other trigger events post-origination. As such, prior empirical work has sought to predict mortgage default using proxies for the portion of homeowners with negative equity or aggregate unemployment rates. While these variables separately measure each side of the double trigger, they provide no information on their intersection; further, those controls were either not available in the public domain during the crisis or lacked predictive power.[6] Our MDRI measure plausibly captures the intersection of the mortgage default double trigger as it tabulates search queries from households considering default who may be concerned about both negative equity *and* a loss of income.

As a proxy for the intersection of the double trigger, one would anticipate that elevated aggregate search for the terms that constitute the MDRI should translate into higher levels of mortgage delinquency and default. From there, recent work on the foreclosure externalities suggests that increases in the MDRI should work to dampen returns to housing (foreclosure neighborhood spillovers). Indeed, Lambie-Hanson (2015) finds that seriously delinquent homeowners neglect housing maintenance, subsequently depressing neighboring property values (Gerardi et al., 2015) and leading to further increases in mortgage defaults (Chan et al., 2013). Anenberg and Kung (2014) also find that higher levels of foreclosures can increase housing supply and thus depress property values.

Using the MDRI, we study the predictive effects of our internet-based mortgage default risk measure on key housing indicators at the daily, weekly, and monthly frequencies.

At the daily level, we examine the predictive relationship between the MDRI and indices that track the costs of subprime mortgage credit-default swaps (contracts that price the risk of default on securitized subprime mortgage bonds traded between parties) via the Markit ABX indices. Intuitively, the ABX represents an index of the cost of default insurance for subprime mortgage backed securities. In this sense, the ABX indices are the financial market analog of the MDRI for subprime borrowers: They reflect market participants' expectations of the risk of subprime mortgage default. Note also that the ABX indices are liquid measures traded frequently by practitioners and institutional investors. Thus, by an efficient markets hypothesis argument, if the MDRI predicts the ABX indices, it would suggest that the household default risk captured by the MDRI was initially unknown to and later discovered by market participants. Indeed, following an increase in the MDRI, financial market participants demand larger premiums to insure subprime mortgage bonds against default. Thus, our index appears to capture key incremental information regarding credit default risk and mortgage market performance. In general, our results regarding the MDRI and subprime credit-default swaps are noteworthy as the ABX indices acted as a leading indicator of the widespread turmoil that permeated through financial markets during the recent crisis (Longstaff, 2010).

At the weekly and monthly frequencies, we assess the predictive effects of the MDRI on a number of house price indices, measures of housing sentiment from household surveys, and delinquency and foreclosure proxies. First, the weekly and monthly results indicate that increases in the MDRI lead to a decrease in house price returns, in line with negative externalities due to delinquency or foreclosure. These results are consistent in- and out-of-sample, across various house price indices including the FHFA, Case-Shiller, and the near real time weekly house price index of Anenberg and Laufer (2014). Further, we find that the MDRI is a leading indicator of housing market sentiment tabulated through the monthly University of Michigan Consumer Sentiment Survey. Specifically, increases in the MDRI are followed by an increase in survey responses that cite that now is a bad time to buy a home because the future is uncertain or because the household cannot afford to buy a home. Hence, the movements in the MDRI reveal mortgage distress before similar information is captured by traditional survey measures. Finally, with regard to delinquencies and foreclosures, we find that the MDRI predicts 60-day delinquencies, a common measure of default in the housing industry.

Use of Google search data to develop a mortgage default risk index has a number of distinct advantages. First, as noted above, our data allow us to directly observe the relative frequency of Google internet search on foreclosure and related terms and thus to infer the extent of mortgage default risk. Other survey-based sentiment indicators, such as those from the University of Michigan, do not ask sensitive questions about mortgage default and it is not clear that respondents would be forthcoming in responding to such questions.[7] 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), whereas traditional housing indices are available only with some lag. Fourth, other proxies such as home foreclosures only provide ex post information on mortgage default decisions, while indicators like the VIX volatility index may reflect economic influences unrelated to housing markets. In sum, our index is useful since it is broad-based, directly captures mortgage default risk, predicts common housing and mortgage variables both in- and out-of-sample, and is available in real time. Given those attributes, the MDRI provides new information and comprises a new and useful tool for market participants, regulators, and policymakers seeking timely insights as to the performance and outlook for housing and related sectors.

[4] It is also not clear that respondents would answer such questions truthfully. Singer (2002) contends that survey respondents are less likely to truthfully answer sensitive questions.

[5] See also Bricker and Bucks (2016). More broadly, the household's overall liquidity position will also include other financial assets and available credit. A decline in available liquidity can therefore also trigger mortgage default (Elul et al., 2010). The borrowers with negative equity may also be at risk of default due to other triggers of mobility, such as divorce. We would like to thank and editor for this latter point.

[6] Proxies of portion of homeowners with negative equity were only available from highly proprietary sources during the crisis (see e.g. Haughwout and Okah (2009)) and only a small percentage homeowners with negative equity default (Foote et al., 2008). Aggregate unemployment rates lack predictive power and are often left out of empirical models of mortgage default (Gyourko and Tracy, 2014).

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

Although the MDRI is to the best of our knowledge the first application of Google data to measure mortgage default risk, other papers have applied search query data in different contexts. For example, Da et al. (2015) use Google data to develop a measure of investor sentiment in equity markets, while Beracha and Wintoki (2013) use Google data to proxy local housing demand.[8] Our work is fundamentally different from previous studies as we use search query data to develop a novel measure of mortgage default risk and show that this measure predicts fluctuations in key housing and mortgage indicators.

The remainder of this paper is organized as follows: Section 2 describes the housing and macroeconomic data; in Section 3, we outline the internet search query used in this paper and the construction of the MDRI; Section 4 discusses the main results; the limitations of the MDRI are outlined in Section 5; and Section 6 concludes.

2. Data

We employ data at the daily, weekly, and monthly frequencies to assess the predictive effects of the MDRI in housing markets. Appendix A provides a complete list of the variables used in this study along with descriptions, data abbreviations, the periodicities for which each variable is used, and data transformations necessary to ensure stationarity. Below in Section 3, we provide a detailed description of the Google internet search query data and the construction of the MDRI.

2.1. Daily data

Our daily dataset includes the Markit ABX indices and the Chicago Mercantile Exchange Case-Shiller (CME CS) futures. The ABX indices are key daily housing market indicators and closely followed by institutional traders and other market participants. The ABX indices track the cost of credit-default swaps on subprime mortgage-backed securities of a certain investment grade and fall as the costs to insure mortgage-backed securities rise.[9] Hence, the ABX indices decline 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. The ABX indices are segregated by credit tranche and range from AAA to BBB-. Separate ABX indices were issued in the first half of 2006, the second half of 2006, the first half of 2007 and the second half of 2007. In our analysis below, we use the "on the roll" log first-differenced ABX returns for each credit tranche. For further details on the ABX indices, see Appendix B.

To track house prices at the daily frequency, we employ the CME CS futures. Although trading in the CME CS index is limited, these futures provide the most up-to-date, real-time, proxy for house price expectations in financial markets.[10] For our analysis below, we consider the log first-difference (returns) of the CME CS futures. House price returns computed using CS data do not account for maintenance, taxes, insurance, and the like.

Our daily dataset also includes a number of daily financial and economic indicators that are used as controls. Specifically, we consider the Aruoba et al. (2012) ADS index that tracks business conditions at the daily frequency; the default spread on corporate bonds (BAA minus AAA corporate bond yields; CorpSpread); the VIX stock market fear gauge; S&P500 stock returns (SPY); the Treasury Spread (10-year Treasury minus 2-year Treasury; TreasSpread); and newspaper-based Economic Policy Uncertainty (Uncertainty) from Baker et al. (2013). The ADS captures the state of the business cycle, CorpSpread is a proxy for credit risk, the VIX and SPY measure the dynamics of equity markets; TreasSpread is an indicator for the yield curve; and Uncertainty measures policy uncertainty and media sentiment.

2.2. Weekly data

At the weekly frequency, house price data are from Anenberg and Laufer (AL; 2014).[11] AL use de-listings, rather than sales, to develop "a more timely house price index" that can be computed in near real time. Indeed, AL compare their index to one that uses the typical repeat sales (Case-Shiller (CS)) methodology and find that their de-listings index yields a house price measure that signals house price movements much earlier than an index that uses the CS methodology. Below, we consider both de-listings (AL) and repeat-sales (CS) log first-differenced house prices (returns) at the weekly frequency.[12] This data ranges from June 2008 to October 2012. Other data at the weekly frequency include the aforementioned daily controls aggregated to a weekly periodicity.

2.3. Monthly data

Last, we consider a number of housing and economic proxies at the monthly frequency. The housing data include housing sentiment from the University of Michigan, house price returns from Case-Shiller and the FHFA, and delinquencies and foreclosures.

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 why "now is a bad time to buy a house." We focus on the survey responses that cite that now is a bad time to buy a house because (1) the respondent cannot afford a home (*CantAfford*); or (2) the future is uncertain (*UncertainFuture*).

At the monthly frequency, we test for predictive effects of the MDRI using quality-adjusted, repeat sales house price indices from both Federal Housing Finance Agency (FHFA) and Case-Shiller.[13] Note that the Case-Shiller house prices are calculated using a three month moving average, producing substantial serial correlation.[14] This makes prediction with regard to housing returns difficult after accounting for the autocorrelation. We employ the monthly Case-Shiller and FHFA house price data at the national level and Case-Shiller data at the metropolitan level. The local Case-Shiller data

[8] For other recent applications of search query data, see Baker and Fradkin (2011); Bollen et al. (2011); Castle et al. (2009); Ginsberg et al. (2009) Goel et al. (2010), Mondria and Wu (2010), Schmidt and Vosen (2011), Stephens-Davidowitz (2011), and Carrière-Swallow and Labbé (2011), Da et al. (2011), and Tkacz (2013).

[9] As suggested by the *Wall Street Journal*, the ABX indices "are some of the most closely watched [subprime mortgage] barometers on Wall Street." *The Wall Street Journal*. June 21, 2007. "Index With Odd Name Has Wall Street Glued; Morning ABX.HE Dose." See also "Goldman Pushes Subprime ABX Index as Housing Rebounds: Mortgages." *Bloomberg News*, November 30, 2012.

[10] As the regular, monthly Case-Shiller house price indices are reported with a two month lag, a CME CS futures contract with a settlement date in month t will be based on a the Case-Shiller house price index form month $t-2$.

[11] We would like to thank Elliot Anenberg for providing us with this data.

[12] As above, these computed house price returns do not include maintenance, taxes, or other related factors.

[13] The FHFA and Case-Shiller data differ in a number of ways. First, 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, etc.

[14] Ghysels et al. (2012) find that the AR(1) coefficient for Case-Shiller national housing returns is 0.94; whereas that for FHFA national housing returns is 0.76. See also Case and Shiller (1989) for more about the serial correlation in housing returns.

cover the largest metro housing markets.[15] All house price data are transformed into log first-difference (return) form.

Our data also include first-differenced delinquency and foreclosure indices from Bloomberg that cover 30-, 60-, and 90-day delinquencies and foreclosures for prime, subprime, and all mortgages at the national level for the monthly periodicity. We transform the first-difference of these indices to have variance 1 over our sample period.

In line with the recent mortgage default literature (Aron and Muellbauer, 2016), our data also includes indicators for mortgage default risk including the interest spread between 30-year fixed rate mortgages and the 30-year Treasuries (MortgSpread), the log of the loan-to-price ratio for new mortgages (Loan-to-Price),[16] and the log first difference in the number of adjustable rate mortgage (ARM) applications (ArmApplications) as proxy for private mortgage credit demand. Other studies that aim to predict mortgage default risk also include the household mortgage debt service ratio and proxies for the portion of mortgage borrowers facing negative equity. Unfortunately, the debt service ratio is only available quarterly, while negative equity proxies are only publicly available for post-crisis data.[17] In unreported results, we include the quarterly debt service as a control by filling in the monthly values with the quarterly average; our below results are unchanged.

Finally, our monthly dataset also contains a number of controls that track housing markets and the macroeconomy. These controls include the corporate default spread (CorpSpread); the Housing Affordability Index (Afford); the civilian employment-population ratio (Employment); housing starts (HouseStarts); industrial production (INDPRO); retail sales (RetailSales); the 10-year Treasury minus the 2-year Treasury (TreasSpread); and newspaper-based economic policy uncertainty (Uncertainty).

3. Internet search query data and the MDRI

We develop our Mortgage Default Risk Index (MDRI) using internet search query data from Google Trends.[18] Google is the most popular search engine in the United States. As of December 2013, Google accounted for 66.5% of all US internet searches.[19] Furthermore, according to the Pew Research Center, 92% of online adults use search engines.[20] Hence, internet queries through Google are representative of the US internet population.

Google Trends reports the search frequency for a given search term relative to all other search terms in the form of a Search Volume Index (SVI). Appendix C provides a detailed overview of the Google Trends SVIs.

To construct the MDRI, we begin by considering housing and mortgage related keywords, such as "mortgage" and "foreclosure," in combination with the word "help." The term "help" is the most commonly queried mortgage default signal according to Google Adwords.[21]

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." 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 Trends on the condition that they contain a housing or mortgage keyword and an indicator of distress. This process leads to 11 key mortgage default and related search query terms.[22] All of the search terms contain a housing keyword along with the word "help" or "assistance."

In general, the search keywords used in the construction of the MDRI were queried frequently by internet users. According to Google Adwords, during 2012, well after the housing crisis peaked, the phrases "foreclosure help" and "mortgage help," two search keywords used in the construction of our index, were queried 266,400 and 594,000 times, respectively. This suggests that our mortgage default risk index is the compilation of millions of internet search queries since 2004 and captures an extensively larger sample than the typical consumer survey.[23]

Figs. 1 and 2 show the screenshots from Google Trends for the SVIs for "mortgage assistance" and "government mortgage help" at the monthly periodicity. As seen in the figure, 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 skyrocket at the peak of the housing crisis. Together, these figures document the timely nature of the mortgage and housing search queries and show how they aptly signal mortgage default risk.

We use Google Trends data at the daily and monthly frequencies. Note that there are differences in how Google Trends reports the data at the different frequencies and we must alter our approach for each periodicity to accommodate these differences. Appendix C provides a detailed overview of our approach.

At the daily frequency, the levels MDRI is the percentage growth in the MDRI from the first period (March 3, 2004). Also, to ensure that we are not constrained by Google Trends'privacy filter, we add the terms "foreclosure" and "foreclosures" to our set of search keywords used to construct the daily index.[24] Our list of search terms used at the daily frequency are in the top panel of Table 1.

To build the monthly MDRI, we query Google Trends for the joint SVI for the 11 search terms in the bottom panel of Table 1.[25]

Last, we do not apply any seasonal adjustment to the MDRI to retain the real-time nature of the index and avoid any "look-ahead" bias.[26] In a previous version of the paper, we seasonally adjusted

[15] These housing markets include Atlanta, Boston, Charlotte, Chicago, Cleveland, Dallas, Denver, Detroit, Las Vegas, Los Angeles, Miami, Minneapolis, New York, Phoenix, Portland, San Diego, San Francisco, Seattle, Tampa Bay, and Washington DC.

[16] Previous literature uses loan to value (LTV), but LTV is only available at the quarterly periodicity for US data. Thus, we use the Loan-to-Price ratio.

[17] The debt service ratio is from the Federal Reserve Board, while estimates for the number of homeowners with negative are publicly available from Zillow beginning in 2011 or can be tabulated from proprietary Core Logic Data. For an example of the latter see Bhutta et al. (2010).

[18] https://www.google.com/trends/ .

[19] As measured by Experian, a company that monitors internet trends.

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

[21] 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. 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.

[22] These search terms include foreclosure assistance; foreclosure help; government assistance mortgage; home mortgage assistance; home mortgage help; housing assistance; mortgage assistance program; mortgage assistance; mortgage foreclosure help; mortgage foreclosure; mortgage help.

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

[24] In unreported results, we also build the monthly MDRI using the search terms listed in the top panel of Table 1. The correlation between our monthly MDRIs that use these different search terms is 0.97.

[25] Specifically, the joint SVI is obtained from Google Trends by entering the search terms separated by a plus sign ("+"). To get the monthly SVI for the MDRI we enter the following into Google Trends: "foreclosure assistance+foreclosure help+government assistance mortgage+home mortgage assistance+home mortgage help+housing assistance+mortgage assistance program+mortgage assistance+mortgage foreclosure help+mortgage foreclosure+mortgage help".

[26] We would like to thank and anonymous referee for this point.

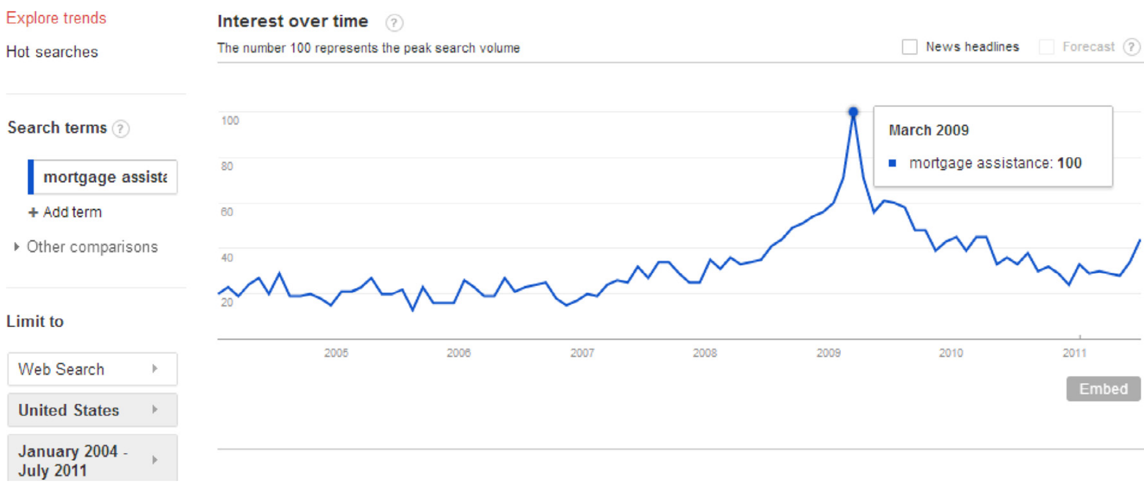

Fig. 1. Google searches for “mortgage assistance”.

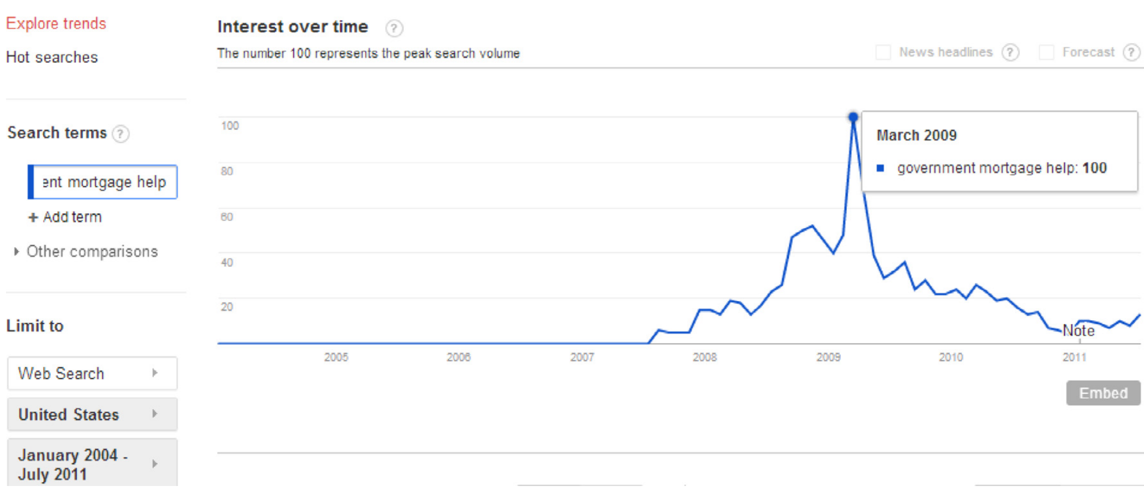

Fig. 2. Google searches for “government mortgage help”.

Table 1
Google Trends search terms.

	Search terms	
Daily	1	foreclosure
	2	foreclosures
	3	foreclosure assistance
	4	foreclosure help
	5	government assistance mortgage
	6	home mortgage assistance
	7	home mortgage help
	8	housing assistance
	9	mortgage assistance program
	10	mortgage assistance
	11	mortgage foreclosure help
	12	mortgage foreclosure
	13	mortgage help
Monthly	1	foreclosure assistance
	2	foreclosure help
	3	government assistance mortgage
	4	home mortgage assistance
	5	home mortgage help
	6	housing assistance
	7	mortgage assistance program
	8	mortgage assistance
	9	mortgage foreclosure help
	10	mortgage foreclosure
	11	mortgage help

the MDRI. The results are similar, indicating that the results presented below are robust to seasonal adjustment. More detail about the MDRI at different periodicities in described in Sections 3.1–3.3, and Appendix C.

3.1. The daily MDRI

We plot the levels daily MDRI versus other key daily housing market and economic indicators in Fig. 3. The shaded bar in the plot represents the bear market over the sample period.[27] As seen in the top panel of the figure, the MDRI stays low in 2004 and 2005, and then rises substantially between 2006 and 2008, highlighting the increase in the mortgage default risk over the sample period.[28] Then in 2009, the MDRI starts to fall as the housing and mortgage crisis begins to abate.

The following three panels show the plots for the ABX AAA, ABX A, and ABX BBB indices. Note that the ABX indices start at 100 on the day of issuance and fall as investors become more bearish on housing and mortgage market performance. While signs of distress surfaced in the MDRI in mid-2006, the highly-rated ABX AAA

[27] Bear markets are defined as a 20% or more drop in the S&P500 over a period of two or more months.

[28] Note also that Permits and housing starts also peaked in late 2005, when the MDRI was lowest. Thus, home builders may have anticipated the housing dynamics reflected in the MDRI. See Liu et al. (2014) and DeCoster and Strange (2012).

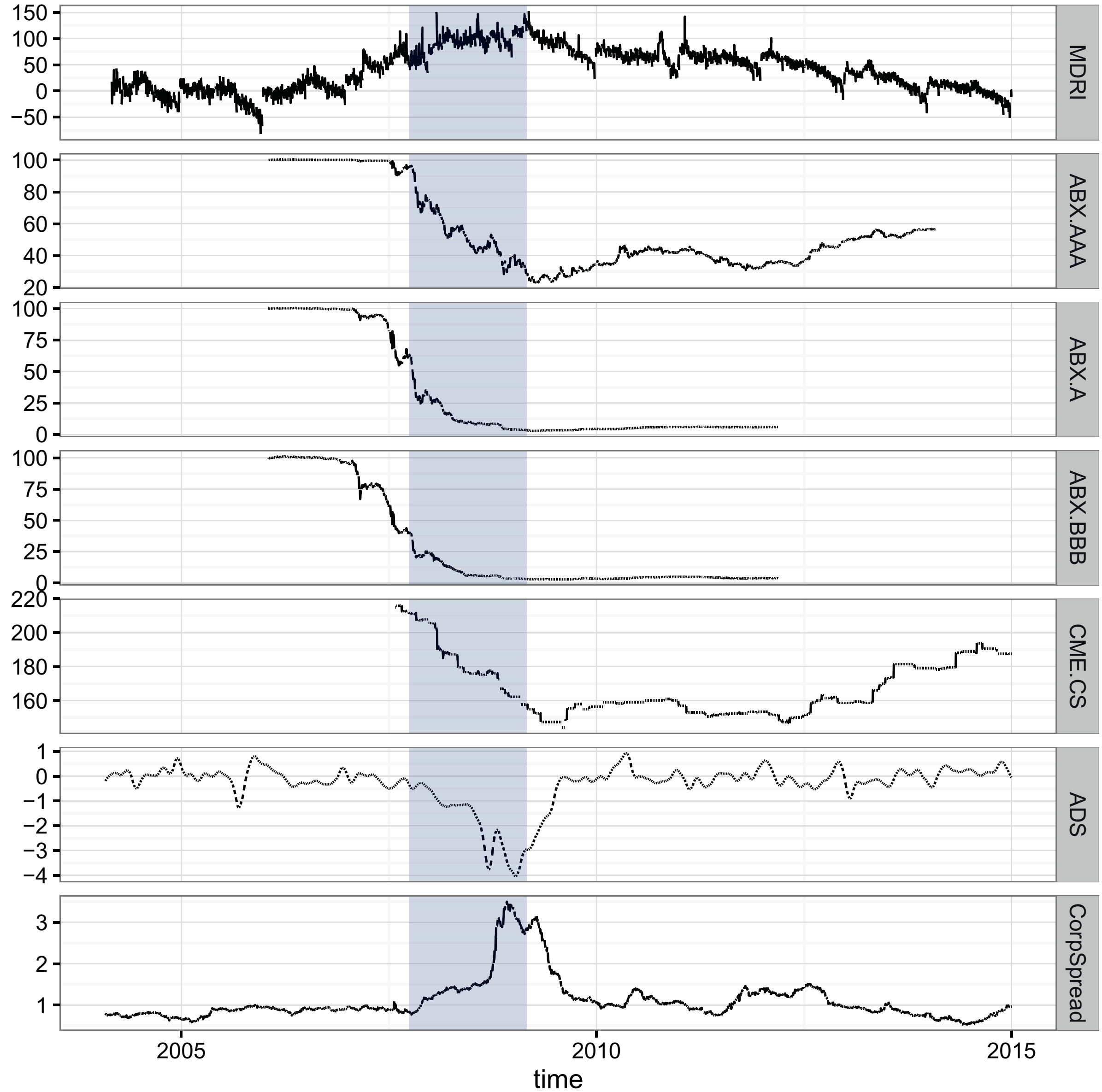

Fig. 3. The daily MDRI and housing market indices.
Notes: Plots of the daily MDRI, the ABX indices that track the cost to insure subprime mortgage backed debt, the CME Case-Shiller futures index, the ADS business conditions index, and the corporate default spread. The shaded bar is a bear market defined as a 20% or more drop in the S&P500 over two or more months.

index did not begin to fall until the start of the bear market in October 2007. From there, the ABX AAA index plunged during the crisis period and then, like the MDRI, recovered in the aftermath of the crisis. The following two panels plot the path of the ABX A and ABX BBB indices. These lower-rated indices also dropped dramatically during the bear market, but never recovered in the aftermath of the crisis.

The next panel shows the path of the CME Case-Shiller futures. Although trading in the CME CS house price futures is limited they closely track the dynamics of the housing market. Indeed, the bottom panel in Fig. 3 shows that the CME CS futures plummeted during the crisis, waffled at the lower end of their range following the crisis and then recovered in the aftermath of the crisis.

Finally, the bottom two panels plot the paths of the ADS business conditions index and the corporate default spread (CorpSpread). These key economic indicators did not reveal signs of crisis until the beginning of 2008, well after the rise in the MDRI.

The correlations between the daily MDRI and the other housing and financial indicators are displayed in Table 2. First, we show the correlation between the MDRI and the VIX index. The VIX index is commonly referred to as a “fear index” for the stock market. So, the large and highly significant correlation coefficient between the daily MDRI and the daily VIX index shown in the table indicates that distress in equity markets was closely related to default risk in housing markets over the sample period. The following rows show the correlations between the daily MDRI and housing indices available at the daily frequency. The results indicate that the MDRI is negatively related to the returns on the CME Case-Shiller (CME CS) futures and ABX indices. These correlation coefficients are all significant at the 10% level.

3.2. The weekly MDRI

We construct our MDRI index at the weekly periodicity for use in the prediction of the near real-time housing price indices of

Table 2
Daily correlations – MDRI, VIX, CME CS, and ABX indices.

	(1) MDRI	(2) VIX	(3) CS CME ret	Subprime Credit Default Swaps (4) ABX AAA ret	(5) ABX AA ret	(6) ABX A ret	(7) ABX BBB ret	(8) ABX BBB- ret
MDRI	1.00***							
	(0.00)							
VIX	0.71***	1.00***						
	(0.00)	(0.00)						
CS CME ret	−0.07***	−0.07***	1.00***					
	(0.00)	(0.00)	(0.00)					
ABX AAA ret	−0.06**	−0.09***	0.05**	1.00***				
	(0.01)	(0.00)	(0.04)	(0.00)				
ABX AA ret	−0.07***	−0.10***	0.00	0.50***	1.00***			
	(0.00)	(0.00)	(0.84)	(0.00)	(0.00)			
ABX A ret	−0.06**	−0.07***	0.04	0.35***	0.65***	1.00***		
	(0.03)	(0.01)	(0.24)	(0.00)	(0.00)	(0.00)		
ABX BBB ret	−0.05**	−0.02	0.02	0.23***	0.46***	0.45***	1.00***	
	(0.04)	(0.50)	(0.58)	(0.00)	(0.00)	(0.00)	(0.00)	
ABX BBB- ret	−0.04*	−0.01	0.03	0.20***	0.42***	0.42***	0.88***	1.00***
	(0.09)	(0.72)	(0.28)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Notes: MDRI is the Mortgage Default Risk Index; VIX is the VIX equity fear gauge; CS CME ret are the returns on the CME Case-Shiller house price futures; and ABX AAA through ABX BBB- returns are the returns on the ABX subprime mortgage credit default swaps. One, two, and three asterisks represents significance at the 10, 5, and 1% levels, respectively.

Anenberg and Laufer (2014). To build our weekly MDRI, we average the daily MDRI over each week. The weekly house price data are available from June 2008 to October 2012.

3.3. The monthly MDRIs

Last, the plots for the national MDRI and other national housing indices at the monthly frequency are displayed in Fig. 4. The top panel of the figure shows the path for the monthly MDRI standardized to have zero mean and unit variance. The MDRI begins to rise in 2006, accelerates upwards in 2007 and 2008, peaks at over 4 monthly standard deviations at the apex of the crisis in early 2009 and then falls in the aftermath of the crisis. Overall, the MDRI appears to closely track mortgage default risk during our sample period.

The plot in the second panel shows the Case-Shiller house price returns. Clearly, these returns document the evolution of the recent crisis: They begin to fall starting in 2006, plunge in 2007, bottom out at around −2% per month in 2008, waffle at around zero percent per month between June 2009 and the end of 2011, and become positive again thereafter. Furthermore, comparing the paths of the MDRI and the Case-Shiller returns suggests that mortgage default risk and housing market returns are highly inversely related. Next, in the third panel of the figure, we plot a University of Michigan survey-based measure of housing market sentiment labelled UncertainFuture. The plot indicates that UncertainFuture did not begin to rise until 2008, well after the start of housing crisis, and remained high after the crisis until 2013. In contrast, the MDRI began to rise in 2006 and fell beginning in 2009 as the effects of the crisis abated. Together, these plots suggest that the MDRI acts as a leading indicator of the survey-based UncertainFuture and thus that mortgage default risk measures a different dimension of household behavior compared to that captured by UncertainFuture.

The final two plots show the path of 30-day delinquencies and foreclosures over our sample period. 30-day delinquencies begin to rise in 2006, jump in 2007 and 2008, and peak in 2009. Hence, the path of delinquencies is similar to that of the MDRI during the lead up to the crisis. Yet unlike the MDRI, 30-day delinquencies only fall slightly in the aftermath of the crisis. This latter result suggests that delinquencies remained elevated in the aftermath of the crisis even as the MDRI began to decline. Finally, the bottom panel of Fig. 4 shows the plot of foreclosures. Foreclosures began to rise in late 2007, shot up between 2008 and 2010, and remained elevated in the wake of the crisis. Comparing the plots of the MDRI and foreclosures suggests that our mortgage default risk proxy acts as a leading indicator of foreclosures over the sample period.[29]

Table 3 displays the correlations between the monthly MDRI and the monthly housing indicators. The results indicate that the MDRI is closely related to housing sentiment measures: The correlation between the MDRI and CantAfford is 0.63, while that between the MDRI and UncertainFuture is 0.68. Thus, the MDRI is related to consumers'concerns about home purchase affordability and future housing and economic uncertainty. The second panel in Table 3 shows the contemporaneous relationship between the MDRI and the monthly US housing returns reported by Case-Shiller and the FHFA. In line with our expectations, the MDRI is highly inversely correlated with housing returns. The following three panels show that the MDRI is correlated with all of the delinquency proxies, but most closely related to instances of delinquencies at 60 days. Indeed, the correlations between the MDRI and the 60-day delinquency measures range from 0.72 to 0.86. These results correspond with our expectations as mortgage lenders and loan servicers typically only seek to work out loan delinquency or initiate foreclosure proceedings after 60 days. Further, delinquency beyond 60 days is considered default in much of the housing literature. See An et al. (2016) for more details.

The bottom panel in Table 3 presents the correlations between the MDRI and the foreclosures. Although the MDRI is correlated with foreclosures, the coefficients are smaller than those found for house price returns or delinquencies. Thus, households appear less likely to signal their risk of default at the foreclosure stage.

[29] As noted by a referee, government programs to limit foreclosures over the sample period may have propped up house prices and pushed foreclosures further into the future. This may explain the elevated levels of delinquencies and foreclosures after the crisis and low levels of housing returns even after the recession ended and the financial crisis abated.

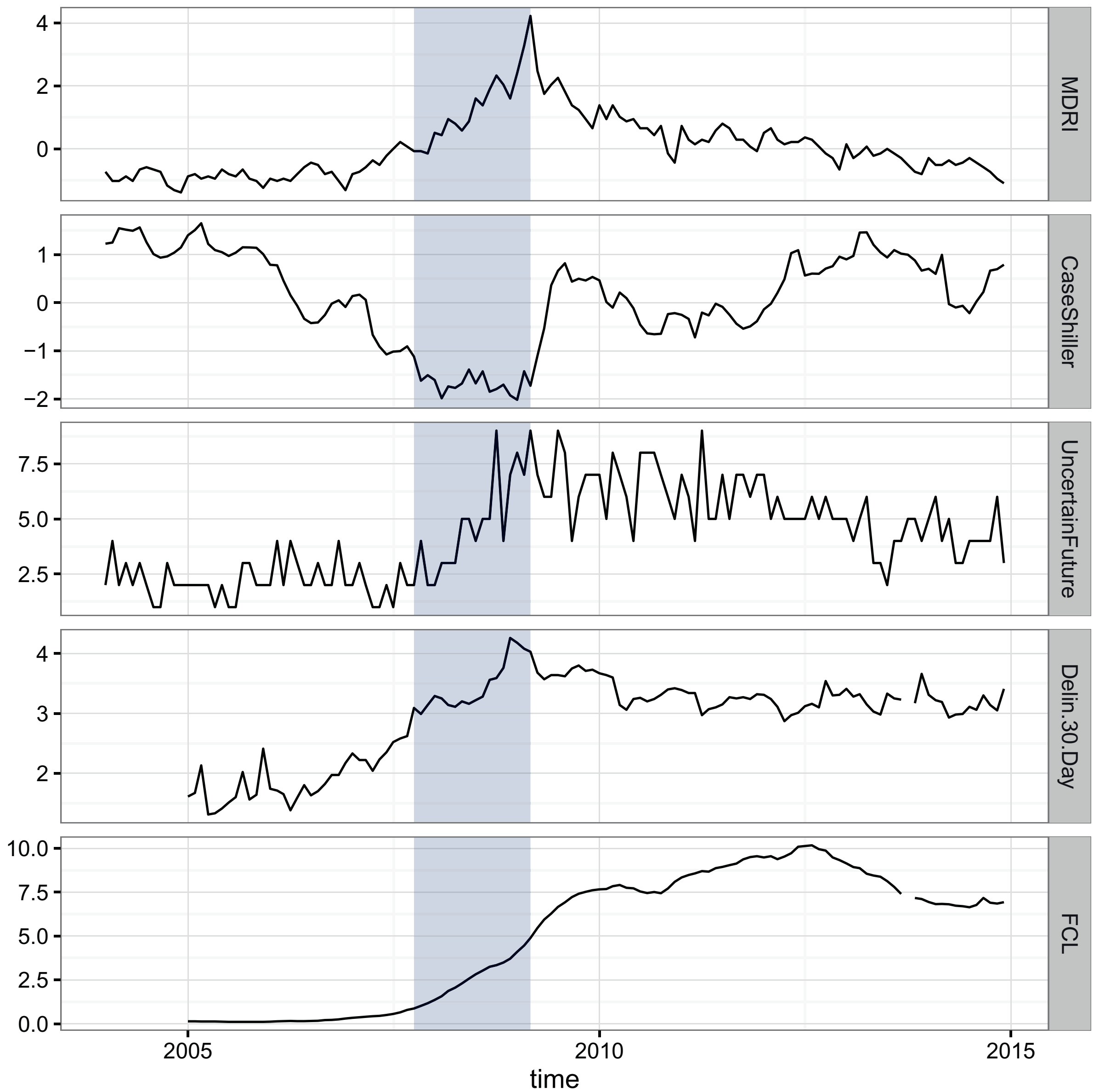

Fig. 4. The monthly MDRI and housing market indices.
Notes: Plots of the standardized monthly MDRI, 20-City Case-Shiller returns, UncertainFuture, 30 day delinquency, and foreclosure indices. The shaded bar is a bear market defined as a 20% or more drop in the S&P500 over two or more months.

Finally, we construct the MDRIs at the local level for the markets where the Case-Shiller house price indices are available. Our methodology is identical to that described above. However, to help circumvent Google Trends'privacy filter and assure adequate local sample size, we also add the words "foreclosure" and "foreclosures" to our list of search terms. We retain the local MDRIs for housing markets where search volume exceeds Google's privacy filter in all time periods. This process yields 16 local MDRIs.[30] We plot the local MDRIs for Dallas, Denver, Miami, and Phoenix in Fig. 5. Specifically, in the plot we show the percentage increase in the local MDRIs from 2006, the peak of the housing boom, to the end of the sample period. Clearly, there is substantial variation in the path of the local MDRIs. For example, there is relatively little movement in the MDRIs for Dallas and Denver, cities that did not experience a major housing boom and bust over our sample period. In contrast, the MDRIs for Miami and Phoenix rise spectacularly during the housing bust beginning in 2007 and then fall markedly in 2011 as the housing market started to recover.

4. Main results

In this section, we assess the predictive effects of the MDRI at the national level, examine the robustness of those results, and study the predictive power of the MDRI across local housing markets. Further, Section 4.2 measures the economic impact of increases in the MDRI on key variables of interest using vector autoregressions; we examine if there is any feedback between the housing market proxies and the MDRI in Sections 4.4; and 4.5 assesses the out-of-sample forecast performance of the MDRI. Note

[30] The local MDRIs are available for Atlanta, Boston, Chicago, Dallas, Denver, Detroit, Los Angeles, Miami, Minneapolis, New York, Phoenix, San Diego, San Francisco, Seattle, Tampa Bay, and Washington DC.

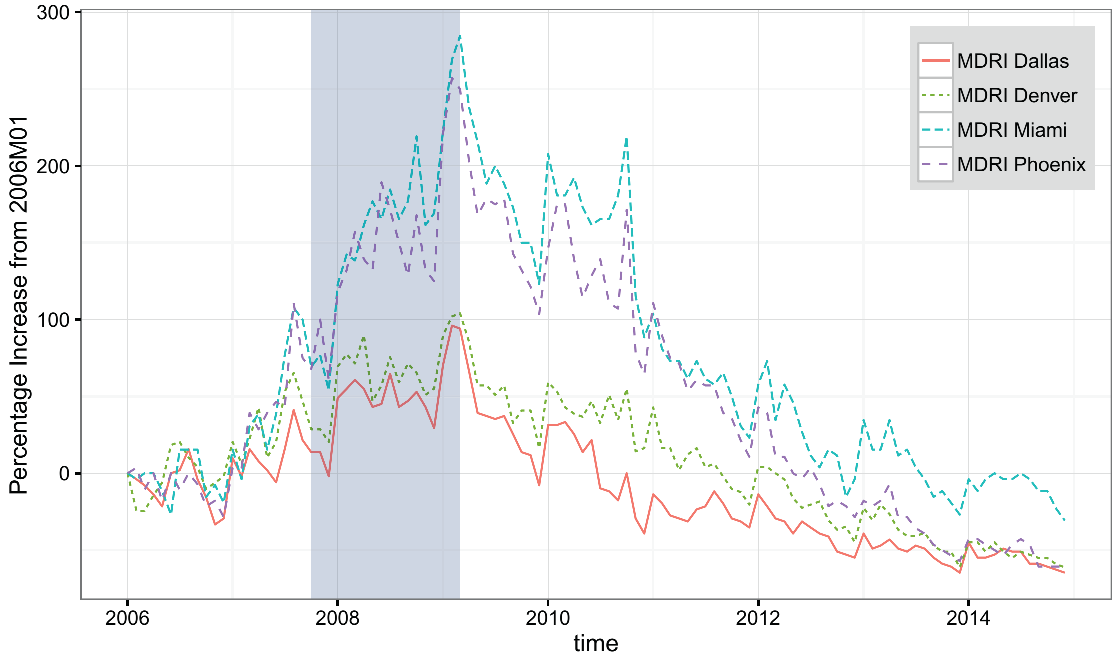

Fig. 5. City-Level MDRIs.
Notes: Plots of cumulative percentage increase in the monthly MDRIs for select cities. The shaded bar is a bear market defined as a 20% or more drop in the S&P500 over two or more months.

Table 3
Monthly correlations – MDRI, housing sentiment, housing returns, delinquencies, and foreclosures.

	MDRI
Michigan Sentiment–CantAfford	0.63***
	(0.00)
Michigan Sentiment–UncertainFuture	0.68***
	(0.00)
CaseShiller US Returns	-0.59***
	(0.00)
FHFA US Returns	-0.63***
	(0.00)
30 Days Delinquent	0.70***
	(0.00)
30 Days Delinquent-Prime	0.49***
	(0.00)
30 Days Delinquent-Subprime	0.68***
	(0.00)
60 Days Delinquent	0.86***
	(0.00)
60 Days Delinquent-Prime	0.72***
	(0.00)
60 Days Delinquent-Subprime	0.85***
	(0.00)
90 Days Delinquent	0.42***
	(0.00)
90 Days Delinquent-Prime	0.36***
	(0.00)
90 Days Delinquent-Subprime	0.50***
	(0.00)
Foreclosures	0.40***
	(0.00)
Foreclosures-Prime	0.29***
	(0.00)
Foreclosures-Subprime	0.51***
	(0.00)

Notes: One, two, and three asterisks represents significance at the 10, 5, and 1% levels, respectively.

that we consider the MDRI minus its 1-year (one-sided) moving average (MDRI_ma) in all of our predictive models.[31] Using the MDRI minus its 1-year moving average allows us to capture the longer term trends in the MDRI while at the same time ensuring stationarity over all sample periods.[32] Note that MDRI_ma is standardized to have zero mean and unit variance. Further, the ABX indices, CME CS futures, and house prices are all in log first-differenced (return) form, and the foreclosure and delinquency measures are in first differences.

4.1. National-level results

We start by analyzing the predictive effects of the national MDRI on national housing market data at the daily, weekly, and monthly periodicities. All lag lengths are chosen by the Akaike Information Criterion (AIC). The robustness of the lag length selection criterion and choice of control variables are assessed in Section 4.3.

4.1.1. The MDRI and daily housing indices

The daily housing proxies considered in this section, such as the ABX indices, are among the most up-to-date, real time indicators of the housing and mortgage market performance. Thus, predictive effects of the MDRI on these daily housing market measures would imply that the MDRI is leading indicator of the US housing market. The results are shown in Table 4. In the top panel, we show the estimated coefficients on lags of MDRI_ma; White heteroskedasticity robust standard errors are in parentheses. The middle panel shows the number of lags chosen for the dependent variable and the exogenous regressors. The control variables include key economic and financial indicators including the ADS business condi-

[31] At the daily frequency, we use the MDRI minus its 250 day (one-sided) moving average. At the monthly frequency, the key measure of interest is the MDRI minus its 12-month moving average.

[32] We would like to thank an anonymous referee for pointing us in this direction.

Table 4
Daily predictive regressions – ABX indices, Case-Shiller futures, the VIX, medical, and the MDRI.

	Returns on Subprime Credit Default Swaps						
	(1) ABX AAA	(2) ABX AA	(3) ABX A	(4) ABX BBB	(5) ABX BBB-	(6) CS CME	(7) ΔVIX
$MDRI\_ma_{t-1}$	−0.11***	−0.04	0.00	−0.07	−0.16**	0.01	0.01
	(0.04)	(0.08)	(0.10)	(0.11)	(0.07)	(0.02)	(0.02)
$MDRI\_ma_{t-2}$		−0.11	−0.15	−0.14		0.02	
		(0.08)	(0.10)	(0.11)		(0.03)	
$MDRI\_ma_{t-3}$						−0.04	
						(0.03)	
$MDRI\_ma_{t-4}$						−0.03	
						(0.02)	
Dep Var Lags	6	6	6	6	6	6	6
ADS Lag	6	5	1	4	1	1	4
CorpSpread Lags	1	1	2	1	1	1	2
ΔVIX lags	6	3	1	1	1	1	NA
SPY Lags	4	3	1	1	1	1	6
TreasSpread Lags	4	3	1	1	1	1	6
Uncertainty Lags	6	6	1	1	1	1	6
MDRI_ma F-stat	9.50***	4.07**	3.01**	4.28**	4.87**	0.90	0.12
R^2	0.12	0.09	0.10	0.02	0.02	0.03	0.04
Adj. R^2	0.10	0.08	0.09	0.01	0.01	0.02	0.03
Start Date	2006-01-20	2006-01-20	2006-01-20	2006-01-19	2006-01-19	2007-08-02	2005-02-24
End Date	2013-12-31	2013-12-31	2012-02-29	2012-02-29	2012-02-29	2013-12-31	2013-12-31
Num. obs.	1971	1969	1517	1517	1515	1597	2207

Notes: Daily predictive regressions of the dependent variable on the MDRI minus its 1-year (250 Day) moving average and controls. The F-statistic in the middle panel tests the null hypothesis that $MDRI\_ma_{t-1} = \cdots = MDRI\_ma_{t-j} = 0$. In columns (1) through (5), the dependent variables are the returns on the ABX subprime credit default swaps; columns (6) and (7) show the results when the Case-Shiller CME housing return futures and the first-difference of the VIX equity fear gauge represent the dependent variable. Control variables include lags of the ADS Business Conditions Index (ADS), the corporate default spread (CorpSpread), the first-difference in the VIX index (ΔVIX), the Treasury Spread (TreasSpread), and Economic Policy Uncertainty (Uncertainty). White heteroskedasticity robust standard errors are in parentheses. One, two, and three asterisks represents significance at the 10, 5, and 1% levels, respectively.

Table 5
Weekly predictive regressions – Case-Shiller and Anenberg-Laufer housing returns and the MDRI.

	Housing Returns	
	(1) CS	(2) AL
$MDRI\_ma_{t-1}$	−0.02*	−0.03
	(0.01)	(0.05)
$MDRI\_ma_{t-2}$		−0.02
		(0.05)
$MDRI\_ma_{t-3}$		0.05
		(0.05)
$MDRI\_ma_{t-4}$		0.11**
		(0.05)
$MDRI\_ma_{t-5}$		−0.10***
		(0.04)
Dep Var Lags	5	5
ADS Lag	4	1
CorpSpread Lags	1	1
ΔVIX lags	1	1
SPY Lags	1	1
TreasSpread Lags	1	1
Uncertainty Lags	1	1
MDRI_ma F-stat	3.83*	2.86**
R^2	0.52	0.33
Adj. R^2	0.49	0.28
Start Date	2008-06-13	2008-06-13
End Date	2012-10-19	2012-10-19
Num. obs.	223	223

Notes: See the notes for Table 4. Weekly predictive regressions of the dependent variable on the weekly MDRI and controls. The weekly MDRI is computed by averaging the daily MDRI each week. CS is the Case-Shiller repeat sales house price index and AL is the Anenberg and Laufer near real time house price index.

tions index, the corporate default spread (CorpSpread), the first-difference of the VIX index (ΔVIX), S&P500 (SPY) stock returns, the 10-year minus 2-year Treasury spread (TreasSpread), and the newspaper-based Uncertainty index. We include the Uncertainty index as the tone of news reports may affect agents'internet search behavior. The middle panel displays the F-statistic corresponding to the Granger Causality test where the null hypothesis is that $MDRI\_ma_{t-1} = \cdots = MDRI\_ma_{t-j} = 0$. The F-statistic is computed using the White heteroskedasticity robust variance-covariance matrix.

The first five columns show the predictive power of the MDRI with respect to the ABX indices that track the cost to insure subprime mortgage backed debt. These results indicate that the MDRI is in fact a leading indicator of the ABX indices: A one standard deviation increase in $MDRI\_ma_{t-1}$ predicts, for example, a decrease in the daily ABX AAA returns of 0.11% points. This effect is statistically significant at the 1% level. Similarly, when any of the other ABX indices represent the dependent variable, the Granger Causality F-statistics are all large in magnitude and significant at the 5% level. Thus, the MDRI appears to be useful in predicting the ABX returns. Last, columns (6) and (7) show the regressions where the returns on the Case-Shiller (CS) CME Futures or ΔVIX represent the dependent variable. In these cases, we find that the MDRI has no predictive power; yet there is limited trading in the CS CME futures and the VIX index is more related to expected risk in the stock market.

In total, the findings in this section imply that the MDRI is a leading indicator of the most up to date, real time measures of housing and mortgage performance. Thus, the MDRI provides practitioners and policymakers with a key real time tool to assess the current direction of household mortgage performance.

Table 6
Monthly predictive regressions – housing sentiment, house price returns, and the MDRI.

	Michigan Sentiment		Housing Returns			
	(1) ΔCantAfford	(2) ΔUncertainFuture	(3) CaseShiller	(4) CaseShiller	(5) FHFA	(6) FHFA
$MDRI\_ma_{t-1}$	0.40*	0.10	−0.05	−0.04	−0.07***	−0.07***
	(0.22)	(0.20)	(0.03)	(0.03)	(0.02)	(0.02)
$MDRI\_ma_{t-2}$	0.13	−0.35	−0.04	−0.04	0.02	0.04
	(0.25)	(0.24)	(0.04)	(0.04)	(0.03)	(0.03)
$MDRI\_ma_{t-3}$		0.35*	0.10***	0.09***		
		(0.20)	(0.03)	(0.03)		
$MDRI\_ma_{t-1}{*}crisis$				−0.04		−0.01
				(0.07)		(0.04)
$MDRI\_ma_{t-2}{*}crisis$				−0.07		−0.08**
				(0.10)		(0.04)
$MDRI\_ma_{t-3}{*}crisis$				−0.10		
				(0.08)		
Dep Var Lags	7	7	3	3	5	5
Afford Lags	1	6	1	1	1	1
ArmApplications Lags	1	4	1	1	4	4
ΔBBMD30 Lags	3	2	1	1	1	1
ΔBBMD90 Lags	1	3	1	1	5	5
CorpSpread Lags	1	7	2	2	7	7
Employment Lags	1	1	1	1	6	6
HouseStarts Lags	2	3	1	1	3	3
INDPRO Lags	2	1	2	2	5	5
Loan-to-Price Lags	1	1	1	1	1	1
MortgSpread Lags	1	1	1	1	1	1
RetailSales Lags	1	2	1	1	5	5
TreasSpread Lags	5	2	1	1	3	3
Uncertainty Lags	1	1	7	7	3	3
ΔVIX Lags	1	4	3	3	2	2
MDRI_ma F-stat	2.45*	1.25	5.74***	6.85***	7.66***	7.28***
MDRI_ma*crisis F-stat				3.63**		2.96*
R^2	0.56	0.77	0.96	0.96	0.97	0.97
Adj. R^2	0.38	0.58	0.95	0.95	0.95	0.95
Start Date	2005-09-01	2005-09-01	2005-09-01	2005-09-01	2005-09-01	2005-09-01
End Date	2014-12-01	2014-12-01	2014-12-01	2014-12-01	2014-12-01	2014-12-01
Num. obs.	111	111	111	111	111	111

Notes: Monthly predictive regressions of the dependent variable on the MDRI minus its 1-year (12 Month) moving average and controls. Controls include lags of the dependent variable, the Housing Affordability Index (Afford), the growth in ARM Applications (ArmApplications), the first-difference on 30 and 90 day delinquencies (ΔBBMD30; ΔBBMD90), the corporate default spread (CorpSpread), the employment-population ratio (Employment), the log of housing starts (HouseStarts), the growth in industrial production (INDPRO), the log of the Loan-to-Price ratio (Loan-to-Price), the fixed rate mortgage - 30-year Treasury spread (MortgSpread), the log first-difference of retail sales (RetailSales), the Treasury Spread (TreasSpread), Economic Policy Uncertainty Index (Uncertainty), and the first-difference in the VIX index (ΔVIX). The first F-statistic in the middle panel tests the null hypothesis that $MDRI\_ma_{t-1} = \cdots = MDRI\_ma_{t-j} = 0$. The second F-statistic in the middle panel tests the null hypothesis that $MDRI\_ma_{t-1}{*}crisis = \cdots = MDRI\_ma_{t-j}{*}crisis = 0$. White heteroskedasticity robust standard errors are in parentheses. One, two, and three asterisks represent significance at the 10, 5, and 1% levels, respectively.

4.1.2. The MDRI and weekly housing indices

Next, we examine the predictive power of the MDRI with regard to weekly house price returns computed using the repeat-sales (Case-Shiller; CS) and de-listings (Anenberg-Laufer; AL) methodologies. The results are in Table 5. The format of Table 5 is identical to that used above in Table 4 and the controls used in Table 5 are the daily variables aggregated to the weekly frequency.

The results from Table 5 indicate that the MDRI predicts weekly house price returns using both the repeat-sales (CS) and de-listings (AL) indices. First, for the repeat-sales CS returns in column (1), higher levels of default risk, as measured by the MDRI, predict lower future house price returns. Yet for the AL house price returns in column (2), the predictive effects are only significant at longer lags. This implies that housing information transmits more slowly through the AL index. Finally, as indicated by the F-statistic in the third panel, we reject at the 10% level the Granger Causality null that the MDRI has no predictive power for house price returns when either the CS or AL house price returns are the dependent variable.

4.1.3. The MDRI and monthly housing indices

In Tables 6 and 7, we assess the predictive power of the MDRI at the monthly frequency. The format of these tables is identical to those discussed above. Model controls include key housing market indicators such as the housing affordability index, standard mortgage default proxies, housing starts, and 30- and 90-day delinquencies. We also control for a number of economic indicators including the civilian employment-population ratio, the growth in industrial production, and economic uncertainty.

First, columns (1) and (2) of Table 6 show the predictive relationship between the MDRI and housing sentiment from the University of Michigan Consumer Sentiment Survey. The results indicate that the an increase in the MDRI, relative to its 1-year moving average, leads to an initial increase in the first-difference in UncertainFuture and the first-difference of CantAfford. While different individual lags of MDRI_ma are statistically significant in both models, only the Granger Causality F-statistic in column (1) where ΔCantAfford is the outcome variable is significant at the 10% level. Next, columns (3) through (6) show the predictive effects of the MDRI on the Case-Shiller and FHFA housing returns.

Table 7
Monthly predictive regressions – delinquencies, foreclosures, and the MDRI.

	30 Days Delinquent			60 Days Delinquent			90 Days Delinquent			Foreclosures		
	(1) ΔTotal	(2) ΔPrime	(3) ΔSubprime	(4) ΔTotal	(5) ΔPrime	(6) ΔSubprime	(7) ΔTotal	(8) ΔPrime	(9) ΔSubprime	(10) ΔTotal	(11) ΔPrime	(12) ΔSubprime
MDRI_ma$_{t-1}$	0.05	0.11	−0.01	0.21***	0.20**	0.24**	−0.33***	−0.05	−0.00	0.02	0.01	0.13*
	(0.07)	(0.09)	(0.11)	(0.07)	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)	(0.10)	(0.07)	(0.08)
MDRI_ma$_{t-2}$		−0.16*					0.22**		−0.08	0.15	0.33***	
		(0.09)					(0.10)		(0.11)	(0.11)	(0.11)	
MDRI_ma$_{t-3}$		0.12					0.02		0.10		−0.54***	
		(0.09)					(0.10)		(0.14)		(0.10)	
MDRI_ma$_{t-4}$							−0.40***		0.01		0.27***	
							(0.07)		(0.13)		(0.10)	
MDRI_ma$_{t-5}$							−0.10		−0.14		0.15*	
							(0.09)		(0.09)		(0.08)	
MDRI_ma$_{t-6}$							−0.04		0.19**			
							(0.10)		(0.09)			
MDRI_ma$_{t-7}$							−0.04		0.19**			
							(0.09)		(0.08)			
Dep Var Lags	6	6	7	7	7	7	2	7	7	7	2	2
Afford Lags	1	1	2	1	1	2	2	2	1	1	1	4
ArmApplications Lags	1	1	1	1	1	1	1	1	1	1	1	1
ΔBBMD30 Lags	NA	1	1	1	5	3	1	1	1	1	1	6
ΔBBMD90 Lags	2	1	1	1	3	2	NA	1	6	4	4	6
CorpSpread Lags	3	1	1	4	3	5	7	1	1	1	1	1
Employment Lags	7	6	1	1	2	1	3	4	1	3	2	2
HouseStarts Lags	1	1	1	3	4	3	5	4	1	7	7	5
INDPRO Lags	6	7	2	2	2	2	1	2	1	1	1	1
Loan-to-Price Lags	1	1	1	1	1	1	4	1	4	4	4	7
MortgSpread Lags	1	1	2	1	1	1	2	2	2	1	1	6
RetailSales Lags	7	6	1	2	2	1	4	5	1	1	2	4
TreasSpread Lags	5	1	3	5	6	5	3	3	3	1	2	1
Uncertainty Lags	1	1	1	5	1	1	6	1	4	1	1	7
ΔVIX Lags	2	1	1	2	2	2	6	1	1	1	7	6
MDRI_ma F-stat	0.38	1.19	0.02	8.37***	4.87**	6.95**	6.60***	0.33	5.79***	1.18	8.05***	2.93*
R^2	0.73	0.74	0.31	0.68	0.69	0.55	0.86	0.78	0.69	0.69	0.82	0.90
Adj. R^2	0.54	0.60	0.09	0.48	0.46	0.31	0.72	0.68	0.49	0.54	0.71	0.75
Start Date	2005-09-01	2005-09-01	2005-09-01	2005-09-01	2005-09-01	2005-09-01	2005-09-01	2005-09-01	2005-09-01	2005-09-01	2005-09-01	2005-09-01
End Date	2014-12-01	2014-12-01	2014-12-01	2014-12-01	2014-12-01	2014-12-01	2014-12-01	2014-12-01	2014-12-01	2014-12-01	2014-12-01	2014-12-01
Num. obs.	111	111	111	111	111	111	111	111	111	111	111	111

Notes: See the notes for Table 6.

In columns (4) and (6), we interact MDRI_ma with a crisis indicator that takes a value of 1 during the crisis and zero otherwise (2007M01 - 2009M03).[33] For the full sample, increases in MDRI_ma predict lower house price returns. A one standard deviation increase in MDRI_ma leads to a decrease in the Case-Shiller house price returns of 0.08% points after two months. Note that although MDRI_ma_{t-1} and MDRI_ma_{t-2} are not statically significant, we do reject the null that these two coefficients are jointly equal to zero at the 10% level. This effect then reverses in the third month. Similarly, when house price returns are calculated using the FHFA index, increases in MDRI_ma lead to a decrease in returns the following month. These predictive effects are summarized by the MDRI_ma Granger Causality F-statistic that is significant at the 5% level when either the Case-Shiller or FHFA returns represent the dependent variable. When we interact the MDRI with the crisis indicator in columns (4) and (6), the results show that increases in MDRI_ma predict even lower housing returns during the crisis for both the Case-Shiller and FHFA indices. Indeed, as seen in the second to the bottom panel of the table, the null that tests that $\text{MDRI\_ma}_{t-1}*\text{crisis}= \cdots = \text{MDRI\_ma}_{t-j}*\text{crisis} = 0$ is rejected at the 10% level for both the Case-Shiller and FHFA returns, providing statistical evidence that predictive effects of MDRI_ma are stronger during the crisis period. Section 4.2 provides more detail regarding the magnitude of these effects and their economic significance.

Next, Table 7 shows the predictive power of the MDRI with regard to delinquencies and foreclosures. Specifically, we consider 30-, 60-, and 90-day delinquencies and foreclosures for all loan types, prime loans and subprime loans. The control set is identical to that used in Table 6. Overall, the results indicate that the predictive effects of the MDRI relative to its 1-year moving average are largest for 60-day delinquencies, 90-day delinquencies, and prime and subprime foreclosures: For example, a one standard deviation increase in the MDRI_ma leads to 0.21 standard deviation increase in 60-day Delinquencies and a 0.24 standard deviation increase in 60-day Subprime Delinquencies. Yet we also find that the MDRI predicts 90-day delinquencies for all loans and subprime loans as well as for prime and subprime foreclosures as suggested by the large and highly significant Granger Causality F-statistics. Thus, our results indicate that the MDRI is a key predictor of later stage delinquencies and foreclosures.

In total, the results in this section imply that the MDRI acts as leading indicator of housing sentiment, house price returns, and 60-day delinquencies. Further, for housing returns, we find that the predictive effects of the MDRI are stronger during times of crisis. Together, these findings suggest that that the MDRI captures household mortgage default risk and yields a more timely indicator of housing performance than previously considered housing market proxies.

4.2. Vector autoregressions

To measure the economic magnitude of the predictive effects outlined in the previous section, we compute reduced form impulse responses within a vector autoregression (VAR) framework at both the daily and monthly frequencies. The daily VAR includes the following variables: MDRI_ma, ABX BBB- returns, ABX BBB returns, ABX A returns, ABX AA returns, ABX AAA returns, and S&P500 returns. At the monthly frequency, we similarly use the following variables: MDRI_ma; Case-Shiller and FHFA returns; the first-difference in 30-Day Delinquencies, 60-Day Delinquencies, 90-Day Delinquencies, Foreclosures; and the growth in Industrial Production.[34] At both the daily and monthly levels, VAR lags are selected using the AIC and we trace out dynamic responses following an innovation that increases MDRI_ma by 2 standard deviations. Increases in MDRI_ma of 2 standard deviations or more are common in our sample and occur 65 times in the daily data. 90% bootstrapped confidence intervals for the impulse response functions (IRFs) are computed using a two-step bias corrected stationary block bootstrap procedure as in Politis and Romano (1994) and Kilian (1998).[35] Note that we also apply the Kilian bias correction to the point estimates and that 1000 repetitions are used in the bootstrap procedure. At the daily frequency, we use a block size of 10 days and trace out the IRFs for 15 days. For the monthly data, the block size is 6 months and we trace out the IRFs for 24 months.

Table 8
Daily VAR cumulative impulse responses.

	Day						
	1	2	3	4	5	10	15
ABX BBB-	−0.50	−0.82	−1.23	−1.32	−1.41	−1.95	−2.41
ABX BBB	−0.56	−0.95	−1.39	−1.48	−1.64	−2.11	−2.58
ABX A	0.28	0.10	−0.19	−0.46	−0.77	−1.33	−1.85
ABX AA	0.28	−0.04	−0.16	−0.41	−0.70	−1.03	−1.43
ABX AAA	0.04	−0.07	−0.16	−0.24	−0.32	−0.79	−1.25
SP500	0.22	0.26	0.21	0.08	−0.03	−0.18	−0.36

Notes: See the notes for Fig. 6. Cumulative responses are computed for up to 15 days following a two standard deviation increase in MDRI_ma.

The results are in Figs. 6 and 7 and Tables 8 and 9. The daily IRFs in Fig. 6 show that a 2 standard deviation innovation in MDRI_ma leads a reduction in the returns across the ABX indices. Further, for all of the ABX indices, the upper confidence bound falls below the zero line at some point and thus these IRFs are nonzero and MDRI_ma Granger Causes the set of ABX indices within this VAR framework.[36] We document the economic significance of dynamic effects in Table 8 and show the cumulative VAR IRFs up to 15 days. The results are telling: After 15 days, a 2 standard deviation innovation in MDRI_ma leads to a 2.41% drop in the ABX BBB-, a 2.58% drop in the ABX BBB index, and a drop of at least 1.25 in the other ABX indices. For the ABX BBB- index, the 2.41% drop corresponds to an increase in the cost to insure $10,000,000 million of BBB- subprime mortgage backed debt of $241,000.[37]

The monthly reduced form IRFs following a two standard deviation innovation that increases MDRI_ma are the black-solid lines plotted in Fig. 7; the corresponding 90% boostrapped confidence intervals are the black-dashed lines. These IRFs indicate that housing returns fall for about three to six months following an innovation in MDRI_ma, while 30-, 60-, and 90-day delinquencies and foreclosures rise. The confidence intervals show that the decrease in housing returns and the increases in 30- and 60-day delinquencies and foreclosures are significant at the 90% level. Thus we reject the null MDRI_ma does not Granger Causes this set variables. The cumulative IRFs in Table 9 show that the effects that follow an increase in MDRI_ma are large in magnitude: After 10 months, Case-Shiller prices fall by 1.89% , FHFA prices fall by 0.61% and 60-day delinquencies rise by 1.96 monthly standard deviations. This

[33] In unreported results, we considered several other dates for the start of the crisis, the results are similar.

[34] Computing structural impulse responses using a Cholesky Decomposition with the variables in the aforementioned order produces similar results.

[35] For other implementations of the bootstrap procedure see Wright (2012) Gabriel and Lutz (2014).

[36] Lütkepohl (2005); p. 54.

[37] This is assuming that the "factor," the portion on mortgages in the pool still outstanding, is equal to 1. See Appendix B for more details.

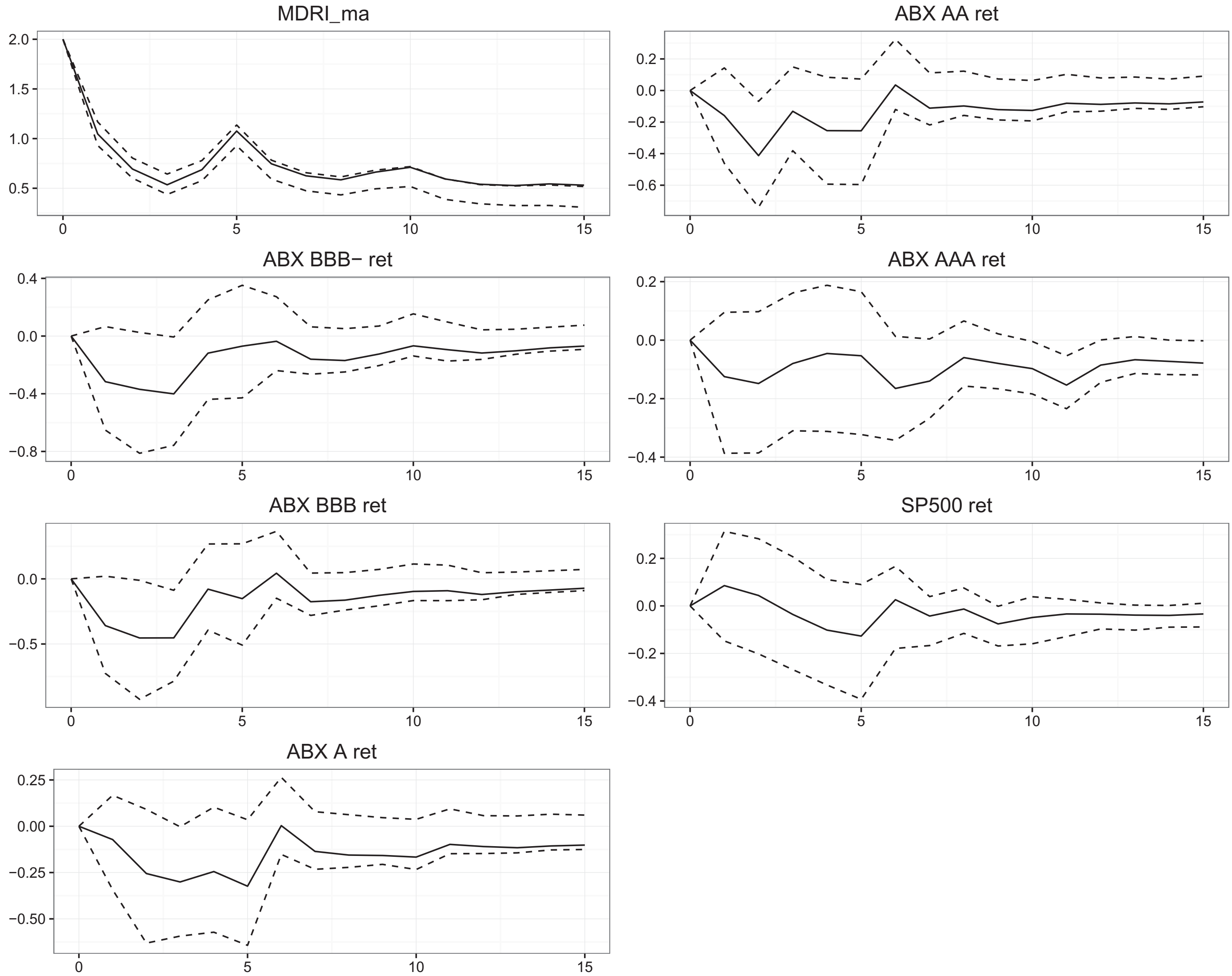

Fig. 6. VAR Daily Impulse Response Functions.
Notes: Daily impulse response functions for a VAR with the following variables: MDRI_ma, ABX BBB- returns, ABX BBB returns, ABX A returns, ABX AA returns, ABX AAA returns, and S&P500 returns. The IRFs are traced out for 15 days following a two standard deviation increase in MDRI_ma.

Table 9
Monthly VAR cumulative impulse responses.

	Month						
	1	2	3	4	5	6	10
Case-Shiller	−0.22	−0.54	−0.84	−1.08	−1.28	−1.45	−1.89
FHFA	−0.13	−0.24	−0.25	−0.22	−0.25	−0.32	−0.61
30-Day	0.31	0.54	0.53	0.54	0.60	0.60	0.60
60-Day	0.46	1.16	1.57	1.70	1.80	1.90	1.96
90-Day	−0.00	0.25	0.56	0.89	1.13	1.33	1.87
Foreclosures	−0.14	0.12	0.44	0.82	1.11	1.36	1.94
Industrial Production	−0.15	−0.70	−1.02	−1.14	−1.30	−1.45	−1.76

Notes: See the notes for Fig. 7. Cumulative responses are computed for up to 10 months following a two standard deviation increase in MDRI_ma.

latter result implies that a 2 standard deviation increase in the MDRI_ma leads to an increase in the percentage of loans that are 60-day delinquent of 0.14% points. For further comparison, the red-dotted lines in Fig. 7 show the reduced form IRFs following an innovation to the Case-Shiller returns of −2% points (equivalent to approximately 2 standard deviations).[38] Based on the 90% bootstrapped confidence intervals (not shown), the reduced form IRFs following a 2 percentage point drop in Case-Shiller returns are only

[38] The monthly standard deviation of the Case-Shiller returns over our sample period is 0.97% .

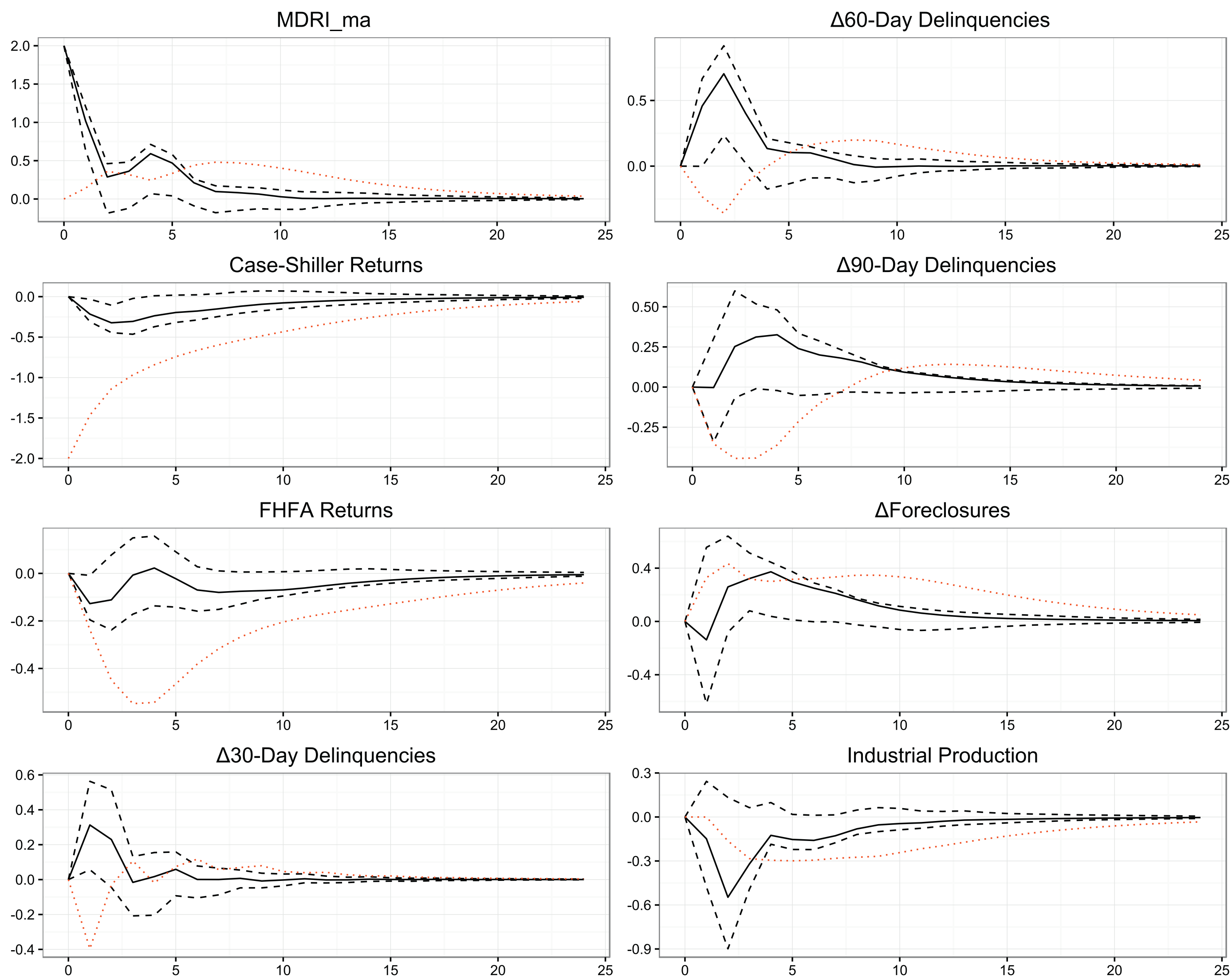

Fig. 7. VAR Monthly Reduced Form Impulse Response Functions.
Notes: Monthly reduced form impulse response for a VAR with the following variables: MDRI_ma; Case-Shiller and FHFA returns; the first-difference in 30-Day delinquencies, 60-Day delinquencies, 90-Day delinquencies, and Foreclosures; and the growth in Industrial Production. The black-solid lines are the reduced form IRFs traced out for 24 months following a two standard deviation increase in MDRI_ma; the black-dashed lines are the corresponding 90% bootstrapped confidence intervals. The red dotted lines are the reduced form responses following a 2% drop in Case-Shiller returns.

significant for FHFA housing returns. Hence, a drop in Case-Shiller returns does not necessarily predict a rise in delinquencies or foreclosures. Moreover, following this decline in house prices, the IRFs for delinquencies initially move in the wrong the direction and are small in magnitude. Together, these results suggest that the MDRI captures the intersection of the income losses and negative equity that constitute the double trigger hypothesis of mortgage default, while housing returns alone have less predictive power.

4.3. Robustness

In this section, we assess the robustness of the foregoing results. Specifically, we consider various permutations of the variables in the control set of our predictive regressions and an alternative lag length selection criterion, the Bayesian Information Criterion (BIC). The full results are in an online appendix. This approach allows us to consider an extensive number of permutations of the control set. We conduct these robustness checks for all of the key dependent variables of interest at the daily, weekly, and monthly frequencies.

We summarize these robustness checks in Table 10. Here, we average the F-statistics across all considered robustness checks and also report the bootstrapped standard errors corresponding to the average F-statistics. In general, the average F-statistics are similar to those found above and thus suggest that our findings are robust to different permutations of the control set and different lag length selection criterion. Further, the bootstrapped standard errors for the average of the F-statistics are generally small in magnitude, implying that there is little variation in the results across model specifications. Overall, the findings from this section indicate that our results are robust to the choice of various control variables and an alternate lag length selection criterion.

4.4. Reverse causality

In addition to examining the predictive effects of the MDRI on key housing indicators, we also consider causality in the other direction: Do available housing market indicators predict the MDRI? Specifically, we build predictive models where MDRI_ma is the dependent variable and the housing market proxies described above

Table 10
Robustness–average F-statistics from predictive regressions.

	F-Stat Mean	F-Stat Mean Boot SE
Daily		
ABX AAA Subprime CDS ret	10.95	0.30
ABX AA Subprime CDS ret	5.84	0.20
ABX A Subprime CDS ret	5.15	0.21
ABX BBB Subprime CDS ret	9.23	0.49
ABX BBB- Subprime CDS ret	9.24	0.37
CS CME Housing Futures ret	1.75	0.13
ΔVIX	0.67	0.08
Weekly		
CS Housing Returns	6.01	0.27
AL Housing Returns	1.74	0.18
Monthly		
Michigan Sentiment–ΔCantAfford	1.87	0.06
Michigan Sentiment–ΔUncertainFuture	1.43	0.07
CaseShiller US Returns	2.34	0.15
FHFA US Returns	11.05	0.20
30 Days Delinquent	0.76	0.05
30 Days Delinquent–Prime	1.20	0.08
30 Days Delinquent–Subprime	0.97	0.05
60 Days Delinquent	4.92	0.20
60 Days Delinquent–Prime	3.50	0.15
60 Days Delinquent–Subprime	3.05	0.15
90 Days Delinquent	2.10	0.12
90 Days Delinquent–Prime	1.30	0.10
90 Days Delinquent–Subprime	2.08	0.13
Foreclosures	1.91	0.14
Foreclosures–Prime	2.43	0.15
Foreclosures–Subprime	1.95	0.15

Notes: The mean of the F-statistics from the regressions in the online appendix. The left column holds the mean F-statistic and the right column shows the bootstrapped standard error of the average F-statistic.

Table 11
Reverse causality – MDRI_ma and predictor variables.

	No Controls	With Controls
Daily		
ABX AAA Subprime CDS ret	2.08	3.54**
ABX AA Subprime CDS ret	1.48	1.53
ABX A Subprime CDS ret	1.96*	2.10*
ABX BBBSubprime CDS ret	1.65	1.70
ABX BBB- Subprime CDS ret	1.23	1.56
CME CS Housing Futures ret	0.22	0.05
Weekly		
CS Housing Returns	2.67*	1.80
AL Housing Returns	0.55	0.59
Monthly		
Michigan Sentiment–ΔCantAfford	0.03	0.15
Michigan Sentiment–ΔUncertainFuture	1.58	5.46***
CaseShiller Returns	6.68**	14.81***
FHFA Returns	1.82	3.32***
30 Days Delinquent	2.96**	1.89*
30 Days Delinquent-Prime	0.27	1.78
30 Days Delinquent-Subprime	1.40	2.01*
60 Days Delinquent	4.16***	3.33***
60 Days Delinquent-Prime	3.71***	4.30***
60 Days Delinquent-Subprime	2.40**	3.03**
90 Days Delinquent	0.17	1.45
90 Days Delinquent-Prime	0.21	0.45
90 Days Delinquent-Subprime	0.30	1.37
Foreclosures	3.47*	10.93***
Foreclosures-Prime	4.40**	9.52***
Foreclosures-Subprime	2.40	1.24

Notes: F-statistics from predictive regressions of MDRI_ma on daily, weekly, and monthly variables. The middle column shows the results when only lags of the dependent variable are used as controls. The right column shows the results when exogenous controls are included in the regression. The controls are the same as those used in Tables 4–6. Lags are chosen using the Akaike information criterion (AIC). One, two, and three asterisks represent significance at the 10, 5, and 1% levels, respectively. Standard errors for the F-statistics are computed using the White heteroskedasticity robust variance-covariance matrix.

represent the key independent variables. Table 11 presents the results with and without exogenous controls. At the daily, weekly, and monthly frequencies, the controls are identical to those used above.

For the daily data at the 5% level of significance, we find that only the higher rated ABX AAA index returns predict MDRI_ma when control variables are included. At the weekly level, the results indicate that house price returns based on the CS methodology are useful in predicting the MDRI when no controls are included, but only at the 10% level of significance. Similarly, in the bottom panel, results using monthly data imply that the Case-Shiller and FHFA returns are also useful in predicting the MDRI. This finding is in line with our expectations and suggests that changes in housing market performance re-enforce the default risk captured by the MDRI.

The F-statistics in Table 11 are significant at the 5% level with and without controls when the 60-day delinquencies serve as the outcome variable. This result is also not surprising as mortgage lenders and servicers usually send notices to homeowners regarding delinquency after 60 days. Thus, upon receiving these notices, we would expect households to search for information about delinquency or default. This further validates the MDRI as a measure of mortgage default risk. Last, we also find that total and prime foreclosures have predictive power with regard to MDRI_ma.

4.5. Out-of-sample forecasts

In the previous sections, we found that the MDRI was a leading indicator of several important housing proxies. Here, we assess the ability of the MDRI to forecast the housing variables of interest out-of-sample (OOS), one-step ahead at the daily and weekly frequencies. Our benchmark model is an autoregressive (AR) model and the forecast model of interest is a linear model with autoregressive lags and lags of the MDRI:

$$r_t = \alpha + \sum_{i=0}^{p} \beta_i r_{t-p} + \sum_{j=1}^{q} \gamma_q \cdot \mathrm{MDRI\_ma}_{t-q} \tag{1}$$

where r_t is the outcome variable of interest and $\beta_0 = 0$, meaning that the selection criterion can choose zero lags of the dependent variable (a random-walk plus drift model). Lags are chosen based on the AIC. Using the BIC for lag length selection produces similar results as shown in an online appendix. Forecasting models are estimated using a number of rolling training windows. Shorter training windows allow the forecasting model to more quickly adjust to changing dynamics in the variables of interest. This may be important given the volatility of financial markets and the macroeconomy over our sample period. Longer windows use more information and may yield more precise parameter estimates.

We assess forecast performance relative to the benchmark AR model using the Diebold and Mariano (2002) test (null hypothesis is equal forecast accuracy between the model and the benchmark) and the OOS R^2 as in Campbell and Thompson (2007) and Goyal and Welch (2008). The lags for the benchmark AR model are chosen using the AIC. The formula for the OOS R^2 statistic is

$$\text{OOS } R^2 = 1 - \frac{MSE_{model}}{MSE_{benchmark}} \tag{2}$$

where MSE_{model} and $MSE_{benchmark}$ are the mean-squared forecast errors for the model of interest and the benchmark AR model, respectively. Positive OOS R^2 values indicate that the model of interest outperformed the benchmark in terms of mean-squared prediction error.

Table 12
Daily one-step-ahead out-of-sample forecast results.

			Returns on Subprime Credit Default Swaps				
(1) Window	(2) Start	(3) End	(4) ABX AAA	(5) ABX AA	(6) ABX A	(7) ABX BBB	(8) ABX BBB-
30	2006-03-06	2012-02-29	4.65	5.95**	6.28**	14.50***	14.92***
			(0.13)	(0.03)	(0.04)	(0.00)	(0.00)
40	2006-03-20	2012-02-29	3.54	3.63	6.48**	10.58***	9.58***
			(0.13)	(0.12)	(0.02)	(0.00)	(0.00)
60	2006-04-18	2012-02-29	2.11	3.78**	4.26**	6.74***	5.76***
			(0.12)	(0.02)	(0.04)	(0.00)	(0.00)
80	2006-05-16	2012-02-29	2.29**	4.31***	3.10	4.59***	5.60***
			(0.04)	(0.00)	(0.11)	(0.00)	(0.00)
100	2006-06-14	2012-02-29	1.41	2.11*	3.01	4.09***	3.70**
			(0.14)	(0.06)	(0.10)	(0.00)	(0.01)
120	2006-07-13	2012-02-29	1.88**	2.07*	1.66	3.94***	3.90***
			(0.04)	(0.07)	(0.26)	(0.00)	(0.00)
140	2006-08-10	2012-02-29	1.26	1.71*	0.89	2.92**	2.92***
			(0.17)	(0.10)	(0.46)	(0.01)	(0.01)

Notes: The daily one-step-ahead out-of-sample R^2 statistics in percentage form. Column (1) shows the training window (in days) and the next two columns show the starting and ending dates of the out-of-sample period for each forecast window. p-values from the Diebold–Mariano test are in parentheses where the null hypothesis is that of equal forecast performance. One, two, and three asterisks represent p-values for the Diebold–Mariano test that are less than 0.1, 0.05, and 0.01, respectively.

We show the daily results for the ABX returns across the various training windows (in days) in Table 12. Columns 2 and 3 of the table display the first and last forecast date for each training window. The remaining columns hold the OOS R^2 statistics (in percentages) for the returns on each of the ABX indices. The p-value from the Diebold–Mariano statistic is in parentheses and one, two, and three asterisks indicate Diebold–Mariano p-values that are less than 0.1, 0.05, and 0.01, respectively. As seen in the table, all of the OOS R^2 statistics are positive and large in magnitude, especially compared to the values found in the equity prediction literature.[39] Further, we reject the Diebold–Mariano null of equal forecast accuracy at the 5% level whenever the ABX BBB or ABX BBB- returns are the outcome variable. As the lower rated ABX indices were key leading indicators of other housing and financial variables during the crisis (Longstaff (2011)), the foregoing results imply that households signalled their default risk via the MDRI before this information was discovered by market participants. The predictive effects are also particularly strong for training windows of 30, 40, 60, or 80 days. The notable performance of the forecasting model at these shorter training lengths suggests that allowing the model parameters to change quickly improves OOS forecast accuracy. This result is not surprising given the volatility of housing markets during the sample period. In an online appendix, we show that our findings are similar when the BIC is used for lag length selection or when the forecast window ends on December 31, 2009, around the conclusion of the crisis.

Table 13 presents the weekly OOS results for house prices constructed using the both the CS and AL methodologies. Note that these weekly HPIs run from June 2008 to October 2012. Overall, all of the OOS R^2 statistics are large in magnitude and the Diebold–Mariano p-value is less than 0.01 when house prices are constructed using the repeat-sales CS methodology. Hence, MDRI_ma appears to be a leading predictor of house price returns out-of-sample.

4.6. Mortgage default risk and the cross-section of housing returns

In this section, we examine the predictive effects of the local MDRIs on metropolitan housing markets using monthly data.

[39] See Campbell and Thompson (2007), Goyal and Welch (2008), and Lutz (2015).

Table 13
Weekly one-step-ahead out-of-sample forecast results.

			Housing Returns	
(1) Window	(2) Start	(3) End	(4) CS	(5) AL
			(0.00)	(0.11)
40	2009-03-20	2012-10-19	6.28**	22.69
			(0.02)	(0.16)
50	2009-05-29	2012-10-19	6.22**	16.86*
			(0.03)	(0.09)
60	2009-08-07	2012-10-19	3.27*	11.61
			(0.08)	(0.15)
70	2009-10-16	2012-10-19	2.56	13.54
			(0.14)	(0.12)

Notes: See the notes for table. The training window is listed in weeks. 12.

Specifically, we employ the following model:

$$r_{it} = \beta_1 \text{MDRI\_ma}_{i,t-1} + \beta_2 \text{Controls} + \nu_{it} \quad (3)$$

where r_{it} is the local house price return for city i, $\text{MDRI\_ma}_{i,t-1}$ is the MDRI minus its 1-year moving average for each city i and Controls is a vector of controls. The controls include three lags of the dependent variable to account for the autocorrelation in house price returns, one lag of housing starts to further capture local housing market dynamics, and one lag of the local unemployment rate to measure changes in the local macroeconomy. We estimate the model using both OLS and GMM. For the OLS estimates, we include month and city fixed effects and compute the Driscoll and Kraay (1998) standard errors that accommodate both cross-sectional and serial correlation. The GMM estimator is the “twoways” estimator of Arellano and Bond (1991) that allows for both individual and time fixed effects. Standard errors for the GMM estimates are calculated using the robust covariance matrix proposed by Windmeijer (2005). For these local models, we use the monthly house price data from Case–Shiller.

The results are presented in Table 14. The left panel holds the OLS estimates; in the right panel we show the GMM estimates. Columns (1) and (3) estimate the model outlined in Eq. (3). In columns (2) and (4), we interact $\text{MDRI\_ma}_{i,t-1}$ with a crisis indicator that takes a value of 1 between 2007M01 and 2009M03 and 0 otherwise. Thus, through columns (2) and (4), we allow the predic-

Table 14
Monthly panel data predictive regressions.

	OLS		GMM	
	(1)	(2)	(3)	(4)
$ret_{i,t-1}$	0.62***	0.62***	0.08*	0.10***
	(0.03)	(0.03)	(0.04)	(0.04)
$ret_{i,t-2}$	0.19***	0.19***	0.14***	0.14***
	(0.04)	(0.04)	(0.04)	(0.04)
$ret_{i,t-3}$	−0.12***	−0.12***	−0.08**	−0.08**
	(0.03)	(0.03)	(0.04)	(0.04)
$MDRI\_ma_{i,t-1}$	−0.04	−0.02	−0.09**	−0.05
	(0.03)	(0.03)	(0.05)	(0.05)
$HouseStarts_{i,t-1}$	0.32***	0.32***	0.47***	0.45***
	(0.09)	(0.09)	(0.13)	(0.12)
$Unemp_{i,t-1}$	0.02	0.02	0.12***	0.12***
	(0.03)	(0.03)	(0.03)	(0.03)
$MDRI\_ma_{i,t-1}*crisis$		−0.05		−0.14*
		(0.05)		(0.08)
R^2	0.78	0.78		
Adj. R^2	0.76	0.76		
City FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes		
Time FE			Yes	Yes

Notes: Predictive Panel Data Regressions of the dependent variable on the city-level MDRI minus its 1-year (12 Month) moving average and controls. The left panel shows OLS estimates with Driscoll and Kraay (1998) robust standard errors in parentheses. In the right-panel, we show the Arellano and Bond (1991) GMM estimates; Windmeijer (2005) robust standard errors are in parentheses. One, two, and three asterisks represent significance at the 15, 10, and 5% levels, respectively.

tive effects of the MDRI to vary across crisis and non-crisis periods. First, as seen in the left panel via the OLS estimates, increases in the local MDRI predict a decrease in local housing returns. Indeed, the estimates in column (1) imply that a one standard deviation increase in the local MDRI relative to its 1-year moving average predicts a decrease in metropolitan level housing returns of 0.04% points. Column (2) indicates that predictive relationship between the MDRI and house price returns is largely concentrated during the crisis period: In the crisis period, a one standard deviation increase in $MDRI\_ma_i$ leads to a decrease in housing returns of 0.07% points. Using OLS, however, the coefficient estimates are not statistically significant.

Next, column (3) shows the GMM estimate for the model in Eq. (3). The results suggest that a one standard deviation increase in $MDRI\_ma_i$ leads to a 0.09% point decrease in local housing returns, an estimate that is significant at the 5% level. Last, column (4) shows the GMM estimates when we allow the predictive impact of the MDRI to vary across the crisis and non-crisis periods. Here, the predictive effects of the local MDRIs are stronger during the crisis period.

5. Limitations of the MDRI

While the MDRI does provide policymakers and practitioners with a novel measure of mortgage default risk, there are some limitations to its use and applications. First, the MDRI captures mortgage default risk at the household level and thus does not yield direct information on the health of mortgage lenders or financial institutions. Hence, high levels of default risk revealed through the MDRI may not necessarily signal a financial crisis or an economic downturn if mortgage lenders and financial firms are sufficiently well positioned to handle the corresponding elevated levels of household defaults. Second, the MDRI is compiled from internet search queries via Google Trends. As such, the MDRI may be vulnerable to changes in internet search behavior. For example, if the Google search engine falls out of favor with consumers or new products (e.g. voice search or searches from mobile devices) alter search patterns, the usefulness of the MDRI may be adversely affected. At this point, we have no reason to believe that recent technological advances are changing internet search behavior, but the risk remains. Further, related to the previous point, households appear to be willing to divulge sensitive information in the context of internet searches.[40] Going forward, internet users may alter their search patterns as they become more aware of data retention policies of various internet entities. This may be relevant for the MDRI as searches for mortgage default keywords are sensitive in nature. Third, the MDRI only reveals default risk known to households. Thus, if a negative shock occurs in a different sector of the economy and then subsequently transmits to the housing sector, the MDRI may act as a lagging indicator during the corresponding downturn. Last, the MDRI is, by its construction, a default risk indicator and hence may be of limited use in the identification of a house price boom.

6. Conclusion

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

We use the MDRI to further examine the predictive relationship between default risk and various indicators of housing market performance. At the daily frequency, we find that the MDRI predicts the ABX indices both in- and out-of-sample. These results suggest that mortgage default risk tabulated through internet search queries acts as a leading indicator of the most up-to-date, real-time measures of housing market performance. Specifically, we find that increases in the MDRI relative to its 1-year moving average lead to lower ABX returns. Hence, elevated levels of the MDRI predict higher costs of default risk insurance for subprime borrowers as reflected in the ABX. Using weekly data, we further assess the predictive relationship between the MDRI and house prices and find that increases in default risk predict lower house price returns for a weekly index based on the Anenberg and Laufer (2014) near real time house price index that exploits delisting information. Similarly, at the monthly periodicity, high levels of the MDRI lead to lower future house price returns across both the Case–Shiller and FHFA house price methodologies. Overall, research findings suggest the utility of internet search data in the development of timely indices of mortgage default risk that provide leading housing market information to researchers, practitioners, and policymakers.

Appendix A. Data list

[40] See, for example, Conti and Sobiesk (2007).

Table 15
Data list.

Mnemonic	Short Description (Variable Abbreviation)	Frequency	Transformation	Source
Mortgage Default Risk Index				
NA	US Mortgage Default Risk Index (MDRI; MDRI_ma)	D,W,M	6	GoogleTrends
NA	City-Level MDRI (MDRI; MDRI_ma)	M	6	GoogleTrends
ABX Indices				
ABX.HE AAA	ABX AAA CDSI (ABX AAA)	D	5	Bloomberg
ABX.HE AA	ABX AA CDSI (ABX AA)	D	5	Bloomberg
ABX.HE A	ABX A CDSI (ABX A)	D	5	Bloomberg
ABX.HE BBB	ABX BBB CDSI (ABX BBB)	D	5	Bloomberg
ABX.HE BBB-	ABX BBB- CDSI (ABX BBB-)	D	5	Bloomberg
House Price Data				
CCICS00	CME-S&P/Case-Shiller HPI Continuous Futures (CS CME)	D	5	Datastream
USCSH1**, USCSH2**	US and City Level Case-Shiller House-Price Indices; SA; 2000M01=100 (CS)	M	5	Datastream
NA	FHFA National House Price Index; SA (FHFA)	M	5	FHFA
NA	Weekly HPI Sales data – CS methodology (CS)	W	5	AL
NA	Weekly HPI Dellisting data (AL)	W	5	AL
Foreclosures and Delinquencies				
BBMD30	Bloomberg 30 Day Mortgage Delinquency Index	M	2	Bloomberg
BBMDP30	Bloomberg 30 Day Prime Mortage Delinquency Index	M	2	Bloomberg
BBMDS30	Bloomberg 30 Day Subprime Mortgage Delinquency Index	M	2	Bloomberg
BBMD60	Bloomberg 60 Day Mortage Delinquency Index	M	2	Bloomberg
BBMDP60	Bloomberg 60 Day Prime Mortgage Delinquency Index	M	2	Bloomberg
BBMDS60	Bloomberg 60 Day Subprime Mortgage Delinquency Index	M	2	Bloomberg
BBMD90	Bloomberg 90 Day Mortgage Delinquency Index	M	2	Bloomberg
BBMDP90	Bloomberg 90 Day Prime Mortgage Delinquency Index	M	2	Bloomberg
BBMDS90	Bloomberg 90 Day Subprime Mortgage Delinquency Index	M	2	Bloomberg
BBMDFCL	Bloomberg Mortgage Foreclosure Index	M	2	Bloomberg
BBMDPFCL	Bloomberg Prime Mortgage Foreclosure Index	M	2	Bloomberg
BBMDSFCL	Bloomberg Subprime Mortgage Foreclosure Index	M	2	Bloomberg
Equity Market Data				
U:SPY	SPY – S&P500 ETF (SPY)	D	5	Datastream
CVXS00	VIX Index (dVIX)	D,W,M	2	Datastream
Economic Data				
USBCIND	ADS Business Conditions Index (ADS)	D,W	1	Datastream
USMACAU	Arm Applications (ArmApplications)	M	5	Datastream
EMRATIO	Civilian Employment-Population Ratio (Employment)	M	1	FRED
MOODCAAA	Moody's AAA Corporate Bond Yields	D	1	Bloomberg
MOODCBAA	Moody's BAA Corporate Bond Yields	D	1	Bloomberg
NA	Corporate Default Spread – MOODCAAA - MOODCBAA (CorpSpread)	D,W	1	NA
I02502Y	2-Year US Treasury	D	1	Bloomberg
I02510Y	10-Year US Treasury	D	1	Bloomberg
NA	Treasury Spread – I02510Y - I02502Y (TreasSpread)	D,W	1	NA
USFM30YR	US 30-Year Fixed Rate Mortgage	M	1	Datastream
FRTCM30	US 30-Year Constant Maturity Treasury	M	1	Datastream
NA	MortgSpread (USFM30YR - FRTCM30)	M	1	NA
INDPRO	Industrial Production (INDPRO)	M	5	FRED
HOUST	Housing Starts (HouseStarts)	M	4	FRED
COMPHAI	Housing Affordability (Afford)	M	4	FRED
USM%PCF3A	Loan-To-Price Ration (Loan-To-Price)	M	4	Datastream
RSXFS	Retail Sales (RetailSales)	M	5	FRED
∗UR	City-level Unemployment Rate (Unemp)	M	1	FRED
∗BP1FHSA	City-level Housing Starts (HouseStarts)	M	4	FRED
Media Uncertainty				
USEPUINDXD	Economic Policy Uncertainty	D,W,M	1	FRED
Housing Sentiment				
NA	Bad time to Buy – Can't Afford To buy; Table 42 (CantAfford)	M	2	U of Mich
NA	Bad time to Buy – Uncertain Future; Table 42 (UncertainFuture)	M	2	U of Mich

Notes: **Table continued on next page.** Time series transformations: 1 - no transformation; 2 - first difference; 4 - logarithm; 5 - log first difference; 6 - time series minus its 1-year moving average. Data are from Anenberg and Laufer (2014; AL), Bloomberg, Datastream, the Federal Housing Finance Agency (FHFA), the FRED Economic Database from the Federal Reserve Bank of St. Louis, Google Trends, and the University of Michigan.

Appendix B. The ABX series

In this appendix we briefly describe the ABX indices. Each ABX index tracks the cost to insure a basket of 20 subprime mortgage backed securities, equally weighted.

The ABX indices are split up based on investment quality and time of issuance. The 2006-01 set of AAA and lower-rated ABX indices that we use in this paper is comprised of loans made in the first half of 2006. We can interpret $(100 - ABX)$ as the up-front payment above the coupon required to insure certain mortgage loans.

To exactly understand how the ABX relates to the cost for insurance we first define the following variables:

- The value for the ABX index (*ABX*). The ABX is always 100 on the day of issuance.
- The *Loan*: The amount of mortgage backed securities to be insured.

- The *Coupon*: The annual fixed payment for the insurance, reported in basis points.
- The *Factor*: The proportion of the principal currently outstanding. This equals one on the day of issuance.

Using the above variables we can calculate the cost to insure a given amount of mortgage backed securities:

$$\begin{aligned}\text{Insurance Cost} &= (100 - ABX)\cdot Loan \cdot Factor \\ &\quad + Loan \cdot Factor \cdot Coupon \\ &= (100 - ABX + Coupon)\cdot Loan \cdot Factor \qquad \text{(B.1)}\end{aligned}$$

The derivative of Eq. (B.1) with respect to ABX is negative. Hence, it becomes more costly to insure mortgage backed securities as ABX falls. In other words, the ABX indices fall as investors become more pessimistic about mortgage backed securities. The "on the roll" ABX index returns that we use in this paper refer to the most recent instance of the ABX indices. For further details on the ABX indices see Longstaff (2010). For further details on CDOs over the housing boom and bust see Cordell et al. (2011).

Appendix C. Google Trends and the construction of the MDRI

In this appendix, we provide further details on the Google Trends Search volume indices and the MDRI.

C.1. Appendix: Google trends search volume index (SVI) and the construction of the daily MDRI

Within each sample period, the SVI for each search term is normalized to range from 0 to 100 and so that a value of 100 represents the date that the given search term achieves peak relative search volume. Indeed, the SVIs are normalized so that the SVI for every search term will achieve a value of 100 over a given sample period. Further, note that Google Trends implements a privacy filter and thus only reports the SVI when the number of absolute searches in a given time period are above a certain unknown threshold. If the number of searches does not exceed this threshold, Google Trends reports a value of 0 in the SVI. Hence, zero values in the SVI represent periods when the number of searches does not meet the Google Trends privacy threshold. Last, the Google Trends data represent a sample of overall Google search data. Therefore, as noted by Choi and Varian (2012), the data from Google can "vary a few percent from day to day." We did not find any substantive differences in the Google Trends data that were downloaded on different days.[41]

C.2. Appendix: construction of the daily MDRI

At the daily frequency, Google Trends reports the data only for three months at a time. So, for every month, we download the data for three months and use the only the last month and the last day of the middle month. To make the daily data comparable across months, we then take the log first-difference of the data within each month. After this process, for every month, we have the log first difference in MDRI. Then, to build a levels index, we accumulate the log first-differenced data over the full sample period (cumulative sum the log first differenced data). The cumulative MDRI in levels is then the percentage growth in the MDRI from the first period. Note that we only retain NYSE market days in construction of the daily MDRI index. The levels daily MDRI thus represents the growth in mortgage default risk from the first available data date, March 1, 2004.

[41] As a robustness check, we re-downloaded the data for the monthly MDRI on several different days. The minimum correlation coefficient between the MDRIs constructed using data from different days was 0.98.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.jue.2016.08.004 .

References

An, X., Deng, Y., Gabriel, S.A., 2016. Default Option Exercise over the Financial Crisis and Beyond. Working Paper.
Anenberg, E., Kung, E., 2014. Estimates of the size and source of price declines due to nearby foreclosures. Am. Econ. Rev. 104 (8), 2527–2551.
Anenberg, E., Laufer, S., 2014. A More Timely House Price Index. Working Paper.
Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. Rev.Econ.Stud. 58 (2), 277–297.
Aron, J., Muellbauer, J., 2016. Modelling and forecasting mortgage delinquency and foreclosure in the UK. J. Urban Econ. 94, 32–53.
Aruoba, S.B., Diebold, F.X., Scotti, C., 2012. Real-time measurement of business conditions. J. Bus. Econ. Stat..
Baker, S., Fradkin, A., 2011. What Drives Job Search? Evidence from Google Search Data. Working Paper.
Baker, S.R., Bloom, N., Davis, S.J., 2013. Measuring Economic Policy Uncertainty. Working Paper.
Beracha, E., Wintoki, M.B., 2013. Forecasting residential real estate price changes from online search activity. J. Real Estate Res. 35 (3), 283–312.
Bhutta, N., Shan, H., Dokko, J., 2010. The Depth of Negative Equity and Mortgage Default Decisions.
Bollen, J., Mao, H., Zeng, X.-J., 2011. Twitter mood predicts the stock market. J. Comput. Sci. 2 (1), 1–8.
Bricker, J., Bucks, B., 2016. Negative home equity, economic insecurity, and household mobility over the great recession. J. Urban Econ. 91, 1–12.
Campbell, J.Y., Thompson, S.B., 2007. Predicting excess stock returns out of sample: can anything beat the historical average? Rev. Financ. Stud..
Carrière-Swallow, Y., Labbé, F., 2011. Nowcasting with google trends in an emerging market. In: Journal of Forecasting. Working Papers Central Bank of Chile, p. 588.
Case, K.E., Shiller, R.J., 1989. The efficiency of the market for single-family homes. Am. Econ. Rev. 79 (1), 125–137.
Castle, J.L., Fawcett, N.W.P., Hendry, D.F., 2009. Nowcasting is not just contemporaneous forecasting. Nat. Inst. Econ. Rev. 201 (1), 71–89.
Chan, S., Gedal, M., Been, V., Haughwout, A., 2013. The role of neighborhood characteristics in mortgage default risk: evidence from New York city. J. Hous. Econ. 22 (2), 100–118.
Choi, H., Varian, H., 2012. Predicting the present with google trends. Econ. Rec. 88, 2–9. S1
Conti, G., Sobiesk, E., 2007. An honest man has nothing to fear: user perceptions on web-based information disclosure. In: SOUPS, 7, pp. 112–121.
Cordell, L., Huang, Y., Williams, M., 2011. Collateral Damage: Sizing and Assessing the Subprime cdo Crisis. Working Paper.
Da, Z., Engelberg, J., Gao, P., 2011. In search of attention. J. Finance 66 (5), 1461–1499.
Da, Z., Engelberg, J., Gao, P., 2015. The sum of all fears investor sentiment and asset prices. Rev. Financ. Stud. 28 (1), 1–32.
DeCoster, G.P., Strange, W.C., 2012. Developers, herding, and overbuilding. J. Real Estate Finance Econ. 44 (1–2), 7–35.
Diebold, F., Mariano, R., 2002. Comparing predictive accuracy. J. Bus. Econ. Stat. 20 (1), 134–144.
Driscoll, J.C., Kraay, A.C., 1998. Consistent covariance matrix estimation with spatially dependent panel data. Rev. Econ. Stat. 80 (4), 549–560.
Elul, R., Souleles, N.S., Chomsisengphet, S., Glennon, D., Hunt, R., 2010. What "triggers" mortgage default? Am. Econ. Rev. 100 (2), 490–494.
Foote, C.L., Gerardi, K., Willen, P.S., 2008. Negative equity and foreclosure: theory and evidence. J. Urban Econ. 64 (2), 234–245.
Gabriel, S., Lutz, C., 2014. The Impact of Unconventional Monetary Policy on Real Estate Markets. Working Paper.
Gerardi, K., Rosenblatt, E., Willen, P.S., Yao, V., 2015. Foreclosure externalities: new evidence. J. Urban Econ. 87, 42–56.
Ghysels, E., Plazzi, A., Torous, W., Valkanov, R., 2012. Forecasting real estate prices. Handbook of Economic Forecasting.
Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L., 2009. Detecting influenza epidemics using search engine query data. Nature 457 (7232), 1012–1014.
Goel, S., Hofman, J.M., Lahaie, S., Pennock, D.M., Watts, D.J., 2010. Predicting consumer behavior with web search. Proc. Nat. Acad. Sci. 7 (41), 17486–17490.
Goyal, A., Welch, I., 2008. A comprehensive look at the empirical performance of equity premium prediction. Rev. Financ. Stud. 21 (4), 1455–1508.
Gyourko, J., Tracy, J., 2014. Reconciling theory and empirics on the role of unemployment in mortgage default. J. Urban Econ. 80, 87–96.
Haughwout, A., Okah, E., 2009. Below the line: estimates of negative equity among nonprime mortgage borrowers. Econ. Policy Rev. 15 (1), 32–43.
Kilian, L., 1998. Small-sample confidence intervals for impulse response functions. Rev. Econ. Stat. 80 (2), 218–230.
Lambie-Hanson, L., 2015. When does delinquency result in neglect? mortgage distress and property maintenance. J. Urban Econ. 90, 1–16.
Liu, C.H., Nowak, A., Rosenthal, S., 2014. Bubbles, Post-crash Dynamics, and the Housing Market. NBER working paper.

Longstaff, F.A., 2010. The subprime credit crisis and contagion in financial markets. J. Financ. Econ. 97 (3), 436–450.
Lütkepohl, H., 2005. New Introduction to Multiple Time Series Analysis. Springer Science & Business Media.
Lutz, C., 2015. Two-Stage Model Averaging: An Application to Forecasting Stock Returns. Working Paper.
Mondria, J., Wu, T., Zhang, Y., 2010. The determinants of international investment and attention allocation: using internet search query data. J. Int. Econ. 82 (1), 85–95.
Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. J. Am. Stat. Assoc. 89 (428), 1303–1313.
Schmidt, T., Vosen, S., 2011. Forecasting private consumption: survey-based indicators vs. google trends. J. Forecasting 20 (6), 565–578.
Singer, E., 2002. The Use of Incentives to Reduce Nonresponse in Household Surveys. Survey Nonresponse, 163–177.
Stephens-Davidowitz, S., 2011. The Effects of Racial Animus on Voting: Evidence Using Google Search Data. Working Paper.
Tkacz, G., 2013. Predicting Recessions in Real-time: Mining Google Trends and Electronic Payments Data for Clues. Working Paper.
Windmeijer, F., 2005. A finite sample correction for the variance of linear efficient two-step gmm estimators. J. Econometrics 126 (1), 25–51.
Wright, J.H., 2012. What does monetary policy do to long-term interest rates at the zero lower bound? Econ. J. 122 (564), F447–F466.