Google search intensity, mortgage default and house prices in regional residential markets

WANG, XIANGDONG

Google search intensity, mortgage default and house prices in regional residential markets

Xiangdong Wang
Department of Finance
Durham University Business School
Durham University

This thesis is submitted for the degree of

Doctor of Philosophy

December 2023

Abstract

The internet provides a new way for households to access relevant information, while their online search behaviour may also contain information for their concerns and intentions, or even be used to predict real economic activity. This thesis explores the use of Google search data to predict mortgage default and regional house price dynamics in an empirical macroeconomic framework. The thesis is composed of three independent empirical studies.

The first study examines the dynamic interdependence between mortgage default and house price across different housing market segments, i.e., top-tier vs. bottom-tier houses and recourse states vs. non-recourse states, based on a Panel VAR model. In particular, this study uses the Mortgage Default Risk Index (MDRI) proposed by Chauvet et al. (2016). It captures the intensity of Google search for keywords and phrases such as "mortgage foreclosure" or "foreclosure help" and measures the potential default risk of households. It is shown that shocks to house price returns have a significantly stronger effect on actual foreclosures in non-recourse states than in recourse states. The results suggest that borrowers are financially sophisticated and strategic as they are less likely to default in recourse states. Additionally, the MDRI has a stronger negative impact on top-tier home price returns, while the foreclosure rate of homes more pronouncedly decreases bottom-tier home price returns. These findings hold for the entire sample and recourse states. However, in non-recourse states, the impacts of the MDRI and the HF on bottom- and top-tier house price returns are about the same.

The second study examines the impact of house prices on the foreclosure rates in the local housing market and explores whether the MDRI helps predict future house prices and foreclosures. In particular, this study uses an error correction framework to capture both the long-run equilibrium fundamental component of house prices as well as the short-run dynamics of house prices, including the component of bubbles. It is found that the MDRI shows a negative impact on both components of house prices but, more importantly, a negative impact on foreclosure rates. Furthermore, it is shown that foreclosure rates are negatively affected by the fundamental component of house prices but are not sensitive to their bubble component. This study sheds new light on the predictive power of household sentiment derived from Google searches on prices and foreclosure rates in local housing markets.

The third study recognizes that, by searching online, households are transmitting information to and simultaneously receiving information from the Google Search engine. While they might divulge information about their financial concerns or vulnerability, they are also gathering information and learning through their search behaviour. This chapter aims to examine the comprehensive impact of the disclosure and information-learning effects of online searches on mortgage default. To that end, based on the assumption of different pre-existing knowledge of households, this study defines two kinds of Google search activities of households, i.e., naïve and sophisticated searches, and practically performed by aggregating the search activities for different query terms. It is found that sophisticated search activity has a positive impact on mortgage delinquency but a negative impact on foreclosure starts, while naïve search activity only positively affects foreclosure starts. The results suggest that the Google search activity of households is a combination of information disclosure and information-learning processes. Furthermore, borrowers are more likely to learn from sophisticated online searches, and they can use the information to avoid foreclosure starts.

Declaration

I, Xiangdong Wang, hereby declare that this thesis is my own work except for other references duly acknowledged in the text. The material contained in the thesis has not previously been submitted for any other degree or qualification in this or any other institution.

The first empirical study, i.e., Chapter 2, is based on the co-authored paper with Prof. Damian Damianov and Dr. Cheng Yan. All three authors made great contribution to this project. One version of the paper entitled "Google search and mortgage default in recourse vs. non-recourse states: A Panel VAR model for the U.S." is submitted for publishing and in now under review.

The second empirical study, i.e., Chapter 3, is based on the co-authored paper with Prof. Damian Damianov and Dr. Cheng Yan. All three authors made great contribution to this project. One version of this paper entitled "Google Search Query, Foreclosure and House price" is published in the Journal of Real Estate Finance and Economics. The citation information is as follows:

Damianov, D. S., **Wang, X**., & Yan, C. (2021). Google search queries, foreclosures, and house prices. *Journal of Real Estate Finance and Economics*, 63(2), 177-209.

Acknowledgements

I would like to express my deepest appreciation to my supervisor, Professor Damian Damianov. His unwavering guidance, profound expertise, and consistent encouragement have been instrumental throughout my Ph.D. journey. I'm equally thankful to Dr. Cheng Yan for his invaluable assistance during my Ph.D. application was invaluable. His guidance, both during his tenure at Durham University and after, has been instrumental to my academic journey. I would also like to express my heartfelt gratitude to Dr. Qi Zhang for supervising my Ph.D. study and for his considerate assistance with my job application.

I am very grateful to my parents for their unwavering love and both financial and emotional support. Their backing has been the cornerstone of my journey through my Ph.D. studies. Additionally, I extend heartfelt thanks to my elder uncle, whose inspiration set me on this distinctive path to the UK. May he rest in peace.

I would also like to thank my friend, Dr. Weiqi Hua, who has always encouraged me and stood by me, particularly through the challenging periods of the pandemic.

Last but not least, my sincere thanks go to Professor Gianluca Marcato from the University of Reading and Professor Michael (Jie) Guo from Durham University for their insightful feedback on my thesis.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and motivation

The 2008 financial crisis illustrates the high correlation between different financial sectors of the macro-financial system. It highlighted how an issue in one financial sector can ripple through other interconnected sectors domestically, eventually impacting the entire financial framework. This can then resonate internationally, influencing the global macro economy, especially in today's globalized era (Eichengreen et al., 2012; Mendoza and Quadrini, 2010). The high interdependency between the two financial sectors that gave birth to the 2008 great crisis, i.e., the residential mortgage market and the housing market in the U.S., attracts particular attention from the public. As one of the most crucial relationships between elements from the two markets, the interdependence between mortgage default and house prices has been the focus of both policymakers and academia. For example, Figure 1.1 shows the dynamics of house prices[1] and the 90+ days mortgage delinquency rate[2] in the U.S. at the national level. Intuitively, one can observe a negative correlation between house prices and mortgage delinquency rates in the U.S.

There is no doubt that the relationship cannot be merely generalized as a negative bidirectional relationship, as it is also determined by other economic or non-economic factors, including among other state legislation regarding mortgage default recourse (Demiroglu et al., 2014), the moral reprehensibility problem of mortgage default (Seiler, 2016), and the default disposition options (Biswas et al., 2020). Studies have already provided a comprehensive understanding of the relationship, both theoretically and empirically, addressing a causality from falling house prices to mortgage default increases (see, e.g., Foster and Van Order, 1984; Elul et al., 2010; Ghent and Kudlyak, 2011), and the reverse causality from rising foreclosures to house price decline (see, e.g., Anenberg and Kung, 2014; Mian et al., 2015).

[1] The house price data is downloaded from the website of Zillow: https://www.zillow.com/research/data/.
[2] The data on mortgage delinquency rates comes from the National Delinquency Survey by the Mortgage Bankers Association and is sourced from Bloomberg.

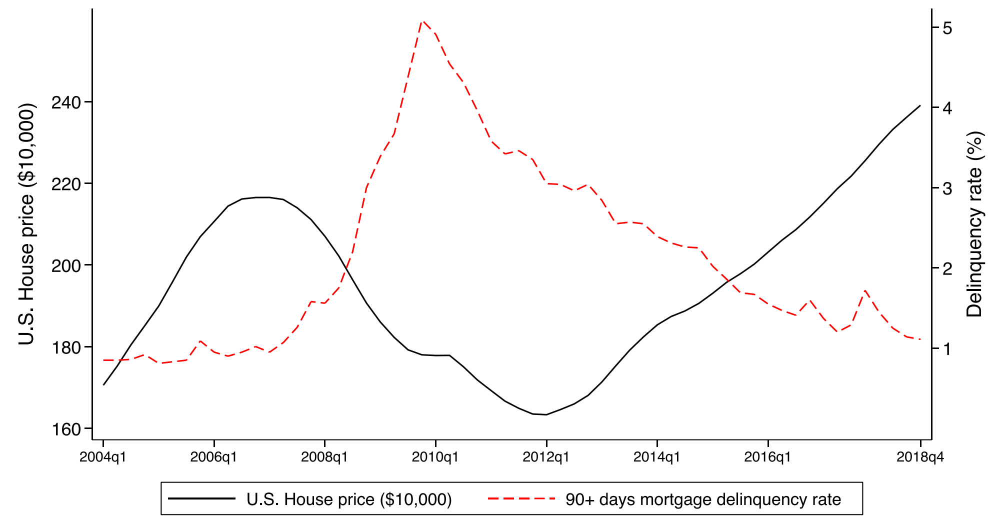


Figure 1.1: U.S. house prices and mortgage delinquency.

Notes: The black solid line represents the dynamics of the house price in the U.S. at the national level, and the red dashed line represents the dynamics of the percentage of mortgages in 90+ days mortgage delinquency in the U.S. at the national level. The house price data is provided by Zillow. The mortgage delinquency rate data is from the National Delinquency Survey conducted by Mortgage Bankers Association, and the data is downloaded from Bloomberg.

However, there are still some research gaps among the current studies on mortgage default and its relationship with house prices. First, a common issue of current studies is that most rely on actual mortgage performance data, such as mortgage delinquency and foreclosure rates. While this ensures the reliability and accuracy of the data, it is highly affected by the time delay in data collection and is therefore not suitable for forecasting purposes.

Second, most studies on this topic utilize micro-level data, i.e., the price of each house and the performance data of each residential mortgage loan. This can help studies take advantage of more observations, controlling house and mortgage loan characteristics. However, it may also make the model quite complex to avoid missing essential control variables and lose the capability to capture market equilibrium effects.

The internet has changed our daily lives and provides a possible solution to the first research gap. Nowadays, it is common for people to search online for relevant information when they want to go out for a trip, find a job, buy a car, and so on. For example, for academic employees in the UK, the website jobs.ac.uk is one of the most important information sources for job vacancies. Although some online searches may arise from the spontaneous urge of the internet user, most of the searches indicate the interest of internet users in a particular topic, such as a trip, a job opening, or a particular brand of car. On this basis, online search data may provide a new predictive tool for actual economic activity. This potential has been explored in various studies in job search, marketing, and healthcare (e.g., Baker and Fradkin, 2017; Yu et al., 2019; Ziebland et al., 2004).

Similarly, when confronted with the possibility of default, households might turn to the Internet for assistance. Therefore, relevant online searches can contain information about their default risk and help predict households' default risk. Figure 1.2 shows the dynamics of the 90+ days mortgage delinquency rate and the Google Search Volume Index (SVI) for the joint query term “foreclosure help + mortgage help”, consisting of two terms commonly used by borrowers in default risk to search for help.[3] It can be seen that the mortgage delinquency rate and SVI show high synchronism. Moreover, the SVI peaks slightly before the mortgage delinquency rate, implying the possibility of using Google search volume data for selected query terms to predict the mortgage default risk of households.

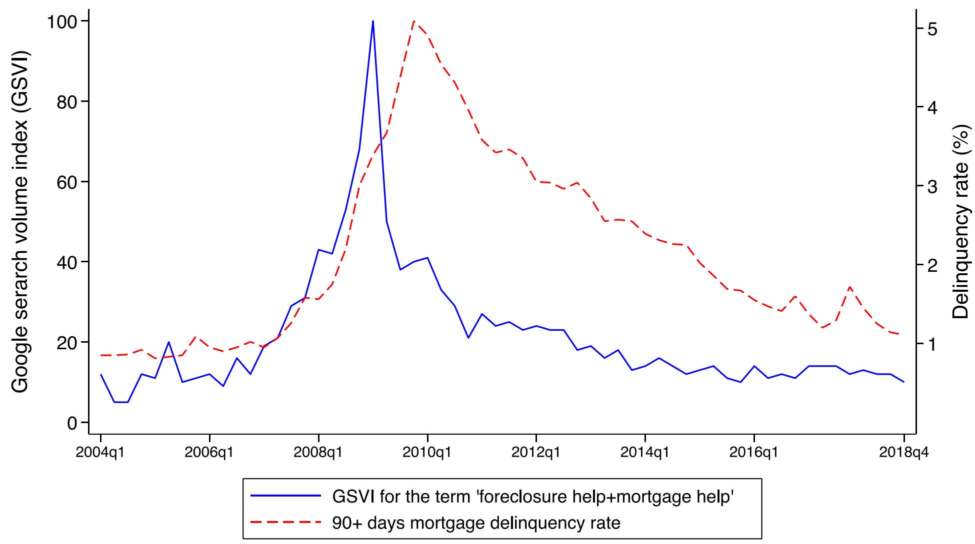


Figure 1.2: Google SVI and mortgage delinquency rate.

Notes: The blue solid line represents the dynamics of the Google Search Volume Index (SVI) for the query term “foreclosure help+mortgage help” for the U.S. at the national level, and the red dashed line represents the dynamics of the percentage of mortgages in 90+ days mortgage delinquency in the U.S. at the national level. The SVI data is provided by Google Trends. The mortgage delinquency rate data is from the National Delinquency Survey conducted by the Mortgage Bankers Association and downloaded from Bloomberg.

Applications of internet search data in the area of real estate markets, particularly to predict the default risk of households, are comparatively less developed, with only a few studies dedicated to this topic (see, e.g., Askitas and Zimmermann, 2011; Chauvet et al., 2016; Webb, 2009). This thesis endeavours to address this gap. The main focus of this thesis is assessing the implication of Google search data, with a particular focus on predicting mortgage default and

[3] The Google Search Volume Index data is provided by Google Trends: https://trends.google.com/trends/.

regional residential house prices. Furthermore, compared with previous studies based on micro-level data, this thesis is conducted within a macroeconomic framework.

1.2 Research questions

To explore the use of Google online search data in predicting mortgage defaults and house prices in an empirical macroeconomic framework, this thesis focuses on the following three groups of research questions:

Question group 1:

- What is the dynamic interdependence between mortgage default and house prices across different housing market segments in an empirical macroeconomic framework?
- Does state-level legislation governing mortgage default have a bearing on the interdependence?

The first empirical chapter, Chapter 2, focuses on the interdependent relationship between mortgage default and house price. To provide a better understanding of the bidirectional relationship, the housing market is disaggregated into different segments along two dimensions: (1) property values, where the housing market is separated into top-tier and bottom-tier according to the house price in each area; and (2) state legislation regarding the right of lenders to pursue a deficiency judgment against borrowers. Based on the categorisation presented by Ghent and Kudlyak (2011) regarding state-level recourse laws, the housing market is separated into recourse and non-recourse states. Furthermore, there are two mortgage default risk measures used in this chapter: (1) the mortgage default risk index (MDRI), which is proposed by Chauvet et al. (2016) based on the Google search activity of households for mortgage default help, is used to measure potential default risk; (2) the percentage of homes foreclosed in each area (HF) is used to measure actual default risk. Based on a Panel VAR model within an empirical macroeconomic framework, it is shown a decline shock in home price returns increases foreclosures more in non-recourse states than in recourse states. The above findings suggest that owners of homes are financially sophisticated and strategic: they are less likely to default in states where lenders can pursue a deficiency judgment against them. Furthermore, the results also show that the MDRI has a stronger negative impact on top-tier homes, while the HF has a stronger negative impact on bottom-tier homes. These findings hold for the entire sample and recourse states. However, in non-recourse states, the impacts of the MDRI and the HF on bottom- and top-tier house price returns are about the same.

Question group 2:

- Does the mortgage default risk index (MDRI) from Chauvet et al. (2016) predict an increase or decrease in future foreclosures and house prices?
- What is the impact of house prices on foreclosure rates? How does it differ between fundamental house prices and bubble components?

The second empirical chapter, Chapter 3, examines the impact of the mortgage default risk index (MDRI), which aggregates the Google search behaviour of households for mortgage default help, on future foreclosures and house prices in local housing markets. Instead of using the MDRI to measure the potential default risk of households, as in Chapter 2, this study focuses more on the possible learning effect of Google searches of households for mortgage help. Furthermore, based on an error correction model accounting for the serial correlation and the mean reversion in the U.S. housing returns time series, we estimate the long-run equilibrium house prices determined by a set of fundamental factors. This helps us to decompose the house prices into their long-term equilibrium component, i.e., fundamental house prices, and short-run deviations from the equilibrium, including their bubble component. The results show that the MDRI negatively affects future foreclosure, supporting that households may learn from their online search. It is also found that the MDRI dampens future fundamental house prices and the bubble component of house prices. Furthermore, the results also show that foreclosure rates increase with the decline of fundamental house prices but are not sensitive to the bubble component of house prices, which aligns with strategic household behaviour.

Question group 3:

- Are households only delivering information about their default risk via their Google search related to mortgage default? Do they learn from their Google searches?
- Will their Google search activities help them avoid mortgage delinquency or keep homes after default?

The final empirical chapter, Chapter 4, focuses on the effect of the Google search behaviour of households for terms related to mortgage default on their default risk, including being in 90+ days of delinquency or falling in the foreclosure process. Theoretically, the online search activity of households can be both information disclosure and learning processes, with the former (latter) process implying a positive (negative) effect on mortgage default. To better understand the comprehensive effects, this chapter examines the effects along two dimensions. First, possessing domain knowledge related to the search topic can determine the choice of query terms and further affect search efficiency. Therefore, based on different assumed possession levels of relevant knowledge households have on the search topic, we define two kinds of search activities, i.e., naïve and sophisticated search activities. Specifically, the former refers to the search activity of households with rare information about how to avoid mortgage default or to keep homes after default, and the latter refers to the search activity of households with enough information on feasible methods. Second, considering that the information-learning process is more time-consuming compared with the information disclosure process, we examine the effects of Google searches on mortgage default in the relatively short term and long term within the last one year. The results show that sophisticated search activity has a positive relationship with mortgage delinquency but a negative relationship with foreclosure starts. In contrast, the naïve search activity is positively related to foreclosure starts. Furthermore, the relationship between Google searches and mortgage default is more significant in states that have experienced a substantial house price drop in the recent four quarters. The results support the idea that the Google search activity of households is a combination of information disclosure and learning processes about their mortgage default risk. They also suggest that borrowers are more likely to learn from sophisticated Google searches to avoid foreclosure starts and keep their houses.

1.3 Contributions

This thesis makes several contributions to the literature regarding the use of Google online search data to predict the dynamics of mortgage defaults and house prices in local housing markets.

Chapter 2 contributes to the literature regarding the interdependent relationship between mortgage defaults and house prices. Unlike most previous studies based on micro-level data, this thesis conducts empirical research based on a Panel VAR model within a macroeconomic framework. Furthermore, compared with a similar study by Calomiris et al. (2013), which is also based on a Panel VAR model, this thesis takes a further step to separate the housing market into different segments from two perspectives, i.e., top-tier vs. bottom-tier housing markets, and recourse vs. non-recourse state markets. In particular, we use the MDRI from Chauvet et al. (2016) derived from Google searches for mortgage help to measure the potential default risk of households, which extends the current study on the use of online search data in the area of real estate. The results provide some findings that align with the strategic default theory. It is also shown that houseowners are less likely to default in states where lenders can obtain deficiency judgment easily.

Chapter 3 contributes to the literature in two ways. (1) It adds to the literature regarding the use of online search data in predicting foreclosures. Unlike previous studies mainly assuming that online search data only provide information about the mortgage default risk of households and imply higher future default risk (see, e.g., Askitas and Zimmermann, 2011; Chauvet et al., 2016; Webb, 2009), this study proposes another two alternative scenarios. First, the online searches are conducted by households faced with mortgage default. However, households may learn from their online searches and adapt their behaviour when dealing with financial distress to avoid foreclosure. Second, the online searches are conducted by internet users for other purposes but not mortgage default, in which case the online search contains no information about the default risk of households. The results suggest that the local foreclosure rates are negatively related to the MDRI, which aggregates online searches for mortgage default help or foreclosure help, supporting the first alternative scenario. (2) While examining the impact of original house price on home foreclosure, this study also uses the methods of Abraham and Hendershott (1996) and Capozza et al. (2004) to separate the original house price into fundamental house price and the bubble component of house prices. The decomposition of house prices enables the analysis and comparison of the impacts of different house price components on foreclosure.

Chapter 4 makes an important contribution to the literature regarding the use of Google search data in the field of real estate research, especially in predicting mortgage default risk. This thesis is the first to analyse the comprehensive impact of Google searches on mortgage delinquency and foreclosure while taking into account the conflicting information disclosure and learning effects of online searches. Previous studies use the data from online searches to construct indicators of mortgage default risk, with the assumption that the search discloses information about the mortgage default risk of households (see, e.g., Chauvet et al., 2016). However, the literature neglects the reverse mechanism that households are likely to learn from their online searches, as documented in the information retrieval literature (e.g., Gadiraju et al., 2018). While information disclosure through online searches of households implies a higher default risk, simultaneous information learning may help them find available options to avoid mortgage delinquency and foreclosure. The two conflicting processes of Google searches make the comprehensive effect of Google searches on mortgage default less predictable. The study in Chapter 4 is the first to find that households are both delivering and collecting information through their online searches regarding mortgage defaults in real estate.

1.4 Structure of the thesis

The remainder of this thesis is structured as follows. The first empirical study is presented in Chapter 2, which examines the dynamic interdependence between mortgage defaults and house prices across different housing market segments. Chapter 3 presents the second empirical study on the predictive power of the MDRI on future house prices and foreclosures. In particular, this chapter pays additional attention to the relationship between the MDRI and the long-term fundamental component and the short-term bubble component of house prices. The third empirical study is presented in Chapter 4, which examines the comprehensive effect of Google searches on mortgage default, with consideration of the conflicting information disclosure and information-learning effects of online searches. The final chapter, Chapter 5, summarises the conclusions, lists some limitations of this thesis, and gives several ideas for future research.

Chapter 2

Google search and mortgage default in recourse vs. non-recourse states: A Panel VAR model for the U.S.

2.1 Introduction

From their peak in March 2007 to their trough in December 2011, top-tier homes in the United States lost, on average, 22 percent of their value as measured by the Zillow home price index. The losses at the low end of the market were even more dramatic as prices of bottom-tier homes declined on average more than 30 percent over the same period. These extreme market convulsions were accompanied by a wave of mortgage defaults, precipitating unprecedented turmoil in financial markets and the most severe economic crisis in recent history. They also stimulated extensive research on the mortgage default behaviour of households and the relationship between foreclosures and house prices.

Most literature on the interdependence between mortgage default and house prices relies on micro-level data. One strand of this literature examines the impact of house prices, foreclosure laws, and other characteristics of local housing markets on mortgage default (e.g., Bajari et al., 2008; Campbell et al., 2011; Elul et al., 2010; Ghent and Kudlyak, 2011; Gerardi et al., 2007). The reverse direction of causality from mortgage default (or other types of financial distress) to house prices has also received considerable attention in the literature (Anenberg and Kung, 2014; Campbell et al., 2011; Gerardi et al., 2015; Lin et al., 2009; Mian et al., 2015).

In contrast to the studies based on household-level data, only some studies attempt to examine the interdependence between mortgage default and house prices in an empirical macroeconomic framework. Foster and Van Order (1984) use aggregate time indices of house prices, unemployment, mortgage rates, as well as household-specific variables (e.g., divorce rate) to show that default probability increases with house price volatility, thus providing empirical support for the model of strategic (or option-based) mortgage default. Calomiris et al. (2013) estimate a vector autoregressive model in which state-level macroeconomic variables such as employment, building permits, and home sales interact with house prices and foreclosures. They find that the effect of house price shocks on foreclosures is substantially greater than that of foreclosure shocks on house prices. That is, the relationship between house prices and mortgage defaults is predominantly based on the endogenous reaction of foreclosures to prices rather than on the downward pressure that foreclosures exert on house prices.[4] These findings support the conjecture that mortgage defaults might occur for strategic reasons, yet the analysis is not sufficiently disaggregated to shed further light on this hypothesis.

In this paper, we undertake such a disaggregated analysis along three dimensions. To begin with, we differentiate between mortgage default risk entailed in the online search behaviour of households (as measured by the indices created by Chauvet et al., 2016) and the actual foreclosures reported by Zillow. Second, we differentiate between recourse and non-recourse states as the extant literature has shown that foreclosure law has an impact on the incentives of borrowers to default on their mortgages (Ghent and Kudlyak, 2011). Finally, we differentiate between top and bottom-tier homes as they exhibit different house price dynamics and are occupied by households with different demographic characteristics.

We use Zillow's tiered home price indices to capture the spatial and market segment differences in house price dynamics. These indices provide relatively broad coverage across U.S. Metropolitan Statistical Areas (MSAs) and account for the differences in appreciation rates between house price segments within the same geographical area. This approach allows us to analyze how the behavioural differences between owners of trade-up and starter homes impact equilibrium outcomes.

To disaggregate mortgage default risk, we use two measures covering the time span from when a household starts seeking mortgage foreclosure help online to when the home is foreclosed. One of the indicators we use is the Mortgage Default Risk Index (MDRI, hereafter) constructed by Chauvet et al. (2016), which measures the default risk of households divulged through online Google searches. The MDRI is derived from the information searching activities by mortgage borrowers and can be perceived as an index measuring the potential default risk of households. Chauvet et al. (2016) find that the MDRI has predictive power for housing returns, mortgage delinquencies, and subprime credit default swaps. The second indicator is the actual number of Homes Foreclosed per 10,000 homes (HF, hereafter) for each

[4] Foreclosures affect house prices through their direct effect on supply or through other types of externalities such as disamenity effects. Calomiris et al. (2013) argue that studies quantifying these effects would help design public policy interventions aimed at breaking the vicious circle of increasing foreclosures and downward spiraling house prices. Based on their analysis, forbearance policies will not be effective in stemming the decline in house prices as the effect of prices on foreclosures is much greater than the effect of foreclosures on prices.

of the studied MSAs. We capture both household sentiment and actual economic activity with these two indicators. Furthermore, disaggregating mortgage default risk and housing market segments helps us better understand the strategic default behaviour at the upper and lower end of the housing market.

To account for the heterogeneity across local residential areas, we estimate a Panel Vector Auto-Regressive (Panel VAR) model (Abrigo and Love, 2016; Glaeser et al., 2014). This macroeconomic equilibrium specification allows us to exploit the variation in local market conditions across different MSAs in order to identify the endogenous relationship between mortgage default risk and house prices without unduly imposing restrictions on the system of equations governing the interdependence among variables (Calomiris et al., 2013; Canova and Ciccarelli, 2013).

We further examine how this relationship depends on foreclosure laws and legal practice across US states. In most states, when a lender forecloses on a home in negative equity, it must obtain a deficiency judgment for the difference between the outstanding mortgage balance and the fair market value[5] of the home (Ghent and Kudlyak, 2011). States differ in their statutes governing the way fair market values are determined, and the time and the costs a lender must incur to obtain a deficiency judgment. These factors determine the extent of recourse of the lender, i.e., the extent to which the lender can collect on mortgage debt beyond the funds raised through the foreclosure proceedings. To investigate this issue, we use the classification of Ghent and Kudlyak (2011), which categorizes states into recourse and non-recourse categories depending on whether pursuing a deficiency judgement is available or practical for lenders.

In line with these results, we find significant differences in the dynamic equilibrium relationships in recourse vis-à-vis non-recourse states. Shocks to house price returns, either for the bottom-tier or top-tier housing market segments, have a significantly stronger effect on actual foreclosures in non-recourse states than in recourse states. These findings are consistent with strategic behaviour: according to the theory of strategic default, homeowners can exercise a put option[6] when their mortgage contracts are non-recourse.

Examining the reverse direction of causality, we find that a shock to the growth rate of the MDRI has a stronger impact on high-value homes, while a shock to the HF has a stronger impact on low-value homes. These findings hold for the entire sample and recourse states. In

[5] The fair market value is determined by a jury or an appraiser, depending on the state's legal practice, and generally does not correspond to the foreclosure resale price.
[6] According to the option theory of default, homeowners prefer to default on their non-recourse mortgage loans when they are sufficiently deep ‘underwater’ even if they can afford their mortgage payments (Kau et al., 1994; Deng et al., 2000).

non-recourse states, the impact of the MDRI and HF on bottom- and top-tier house price returns are about the same.

The remainder of this paper is organized as follows. Section 2.2 reviews the literature on the interaction between house prices and default risk. Section 2.3 introduces the state legislation differences regarding lender recourse across U.S. states and develops our hypothesis. Our sample and methodology are introduced in Section 2.4 and Section 2.5, respectively. The empirical results are presented in Section 2.6, and the concluding remarks in Section 2.7.

2.2 Literature review

Much of the existing literature on the interaction between house prices and the default behaviour of households is based on household-level data, whereby both directions of causality have received considerable attention.

2.2.1 The impact of mortgage default on house prices

It is widely documented that mortgage default hurts house prices. Campbell et al. (2011) estimate a foreclosure discount as high as 27 percent of the average value of a house due to possible damage to the home and the lender's incentive to accept a lower price to sell the house quickly. Mian et al. (2015) use state judicial requirements as an instrument to account for the endogeneity between foreclosures and house prices and find that foreclosures are responsible for about 33% of the decline in house prices during the 2007-2009 period. Recent studies also assessed the negative externality caused by neighbouring foreclosures. Lin et al. (2009) estimate a price impact of 8.7 percent for properties within 100 meters of the foreclosure and 4.7 percent for properties within 200 meters. Campbell et al. (2011) find that forced sales due to foreclosure, bankruptcy, or death of the owner result in about 3%-7% discounts on neighbourhood house prices. A similar negative impact of foreclosure on neighbourhood house prices is also documented by Rogers and Winter (2009). Other studies have also examined the channels through which foreclosures exert downward pressure on prices. Harding et al. (2009) provide evidence of a contagion effect of foreclosure on nearby properties caused by the negative externality of the foreclosed property, which is poorly maintained and neglected in the foreclosure process. In comparison, Hartley (2014) differentiates between a supply effect and a disamenity effect and finds only the former to be significant. Similarly, Anenberg and Kung (2014) disaggregate the effect on prices into a competition and disamenity effect and show that the latter effect is relevant for high-density low-price neighbourhoods. Meanwhile, as foreclosure may cause additional crime in the local area due to poor maintenance (Ellen et al., 2013), it can further increase the negative impact of foreclosure on local house prices.

Alongside foreclosures, alternative methods of property disposition have been increasingly used during the subprime crisis. Clauretie and Daneshvary (2011) explore sales of properties in default through one of the following three options: pre-foreclosure “short sale” by the borrower in default, sales of properties during the foreclosure process by the borrower, or sales of properties as real estate owned (REO) where the lender sells the foreclosed property. They find that the short sale has the lowest price discount but the longest marketing time (see also Biswas et al. 2020). Goodwin and Johnson (2017) find that short sales, similar to REO sales, are sold at a discount, yet short sales are associated with a much longer time on the market.

Another strand of the literature compares the spillover effect of short sales, REO properties, and foreclosed properties and finds supporting evidence for the spillover effects of REO and foreclosed properties but no evidence of negative externality for short sales (Danesshvary et al., 2011; Danesshvary and Claureti, 2012; Depken et al., 2015).

2.2.2 Impact of house price decline on mortgage default

There is a long tradition in the household finance literature for examining the role of price declines and negative home equity on mortgage default. Foster and Van Order (1984) view default as an American put option with an exercise price equal to the value of the mortgage. The decision to default is more complicated when one considers the role of the transaction costs of selling the house and the cost of default in this purely financial model. When these aspects are taken into consideration, equity is still an essential determinant of default, yet households need to be more profound in negative equity territory for strategic default to be advantageous. While Foster and Van Order (1984) do not find empirical support for the purely option-theoretic model, the weaker version of their model that accounts for transaction costs explains the data well.

The empirical literature on strategic default behaviour developed rapidly in the aftermath of the subprime mortgage crisis. Gerardi et al. (2018) find that about 38% of mortgage defaults are caused by strategic motives, which are present when households default, although they could continue to make their mortgage payments without reducing consumption. Much of the recent literature focuses on the “double trigger hypothesis”, which examines the way negative equity interacts with liquidity problems of households (e.g., adverse life events such as bankruptcy, job loss, divorce, illness) to trigger a default (see, e.g., Cunningham et al., 2021;

Elmer and Seelig, 1999; Mocetti and Viviano, 2017; and Tian et al., 2016). Cutts and Merrill (2008) also state that the mortgage default rate cannot be attributed to negative equity alone; rather, income shocks, excessive obligations, and health-related problems are primary causes of serious mortgage delinquency among prime borrowers.

Another way to study strategic default is by examining how lender recourse impacts default behaviour. Ambrose et al. (1997) find that lenders having recourse to assets of borrowers other than the house face a lower incidence of default. Ghent and Kudlyak (2011) show that borrowers in negative equity are more likely to default in non-recourse states, whereby the effect is stronger for owners of high-value homes. Homes in the $500,000-$750,000 value range are twice as likely to default in non-recourse states than in recourse states. Furthermore, Zhu and Pace (2015) find that the expected delay between the first missing mortgage payment and the final foreclosure sale has a significant positive impact on the mortgage default risk of borrowers.

This paper complements these studies by adding household online search behaviour as a measure of mortgage default. We also undertake additional disaggregation across mortgage default risk measures and house price segments. Consistent with previous literature, we find that the interdependence between house prices and mortgage default risk is shaped by the state laws governing foreclosures. However, when we disaggregate the price dynamics, we find that strategic behaviour impacts mainly the upper end of the market.

2.2.3 Dynamic macro-level effects

Only a few attempts have been undertaken in the literature to model the dynamic interdependence between house prices and foreclosures in a system that accounts for the broader macroeconomic environment. Most closely related to the present analysis is the study by Calomiris et al. (2013), who show that the negative impact of prices on foreclosure dominates the impact of foreclosure on house prices. These findings highlight the importance of the strategic choices of homeowners and lenders in shaping these bi-directional dynamics. We further disaggregate house price indices into price tiers and differentiate between mortgage default risk derived from online search behaviour and risk associated with actual defaults. This approach allows us to assess further the occurrence of strategic behaviour in housing and mortgage markets and its impact on equilibrium outcomes. Van Dijk and Francke (2018) have shown that variables related to internet search behaviour, such as the number of clicks on properties listed for sale, are a good proxy for demand and impact prices and liquidity in local residential markets. On the other hand, we demonstrate that internet search behaviour related to mortgage default also impacts house prices yet has a stronger impact on the top segment of the market.

2.3 State legislation difference and hypothesis development

2.3.1 The difference in state legislation: Recourse vs. Non-recourse

A home is in negative equity when the value of the home is lower than the outstanding mortgage balance. Negative equity is one of the triggers of mortgage default according to the double-trigger hypothesis. When a delinquent homeowner is in negative equity, the proceeds from the foreclosure sale will fall short of the outstanding balance. In a foreclosure the gap between outstanding balance and resale price typically widens because many foreclosure sales are fire sales. In other words, even though lenders can recover part of the remaining mortgage balance through the sale of the foreclosed property, they will still suffer losses due to the negative equity of homes.

The difference between recourse and non-recourse states lies in the extent of recourse the lender has to the borrowers' assets beyond the property used as collateral for the initial loan. In order to recover this unsecured debt, lenders need a deficiency judgment or court order. The deficiency judgment covers the difference between the mortgage balance and the fair market value of the foreclosed property rather than the foreclosure sale price (Ghent and Kudlyak, 2011). This is because the lender is often the only bidder at the foreclosure sale and is likely to profit by bidding at a low price without such a restriction.

U.S. states significantly differ in the probability of a lender obtaining a deficiency judgment. Some states, such as Arizona and Oregon, do not allow deficiency judgments. There are also significant differences in the time and cost that lenders must incur to get a deficiency judgment across states allowing such judgment. For example, as the fair market value is usually determined by a jury or an appraiser, there can be additional financial and time costs depending on legal practice in the state. Furthermore, the relevant cost of foreclosure is higher in states with judicial requirements where lenders have to get court approval for foreclosure auctions (Mian et al., 2015). In practice, the deficiency judgment of lenders can also be affected by factors, such as the allowed collection period for lenders on the deficiency after foreclosure sale, types of mortgages that allow a deficiency judgment, and types of assets allowed for deficiency collection (Ghent and Kudlyak, 2011). These factors determine the extent of recourse of the lender, i.e., the extent to which the lender is able to collect on mortgage debt beyond the funds raised through the foreclosure resale proceedings.

2.3.2 Hypothesis development

In recourse states, lenders are more likely to try to recoup the debt not recovered through the foreclosure process by going after other assets of the borrower. Therefore, recourse increases the financial cost of mortgage default for borrowers. Hence, borrowers in recourse states are more concerned about their default risk, prompting them to explore alternatives to prevent it, such as making the required mortgage payments or selling their homes to repay their mortgages before foreclosure. In comparison, in non-recourse states, it is forbidden or impractical for lenders to get a deficiency judgment, and they can only seize and sell the property used as collateral if the borrower defaults. As a result, the financial loss of the delinquent borrowers is restricted to their homes used as collateral, making them less concerned about foreclosure and more prone to default and subsequent foreclosure.

Previous literature has provided evidence that suggests a significant impact of state recourse legislation on the mortgage default risk of households. Ghent and Kudlyak (2011) document that borrowers are more likely to default in non-recourse states, and they prefer to use lender-friendly procedures, such as deed-in-lieu of foreclosure, when they default in recourse states. This implies that borrowers can make strategic decisions about whether to default or avoid foreclosure. Similarly, Demiroglu et al. (2014) find that the default risk is higher for homes with negative equity in states with judicial requirements for foreclosure or in states without deficiency judgment.

With the assumption that borrowers can default strategically, due to the additional default cost related to lender recourse, they are likely to default more in non-recourse states than in recourse states when facing negative equity as a result of drop in house prices. Therefore, we explore the following hypothesis:

Hypothesis 1*: The impact of house price decline on the mortgage default risk of households is more significant in non-recourse states than in recourse states.*

Furthermore, borrowers from different housing market segments may also show different default behavior. As the owners of top-tier homes are usually wealthier than those of bottom-tier homes, they are likely more familiar with the consequences of financial investments and hence more financially literate. Anderson et al. (2022) find that the attitude of borrowers towards strategic default is negatively related to their income and financial experience. Ghent and Kudlyak (2011) also find that the deterrent effect of lender recourse is significant only for high value homes. Building on the prior hypothesis that the impact of house price decline on the mortgage default risk of households is more significant in non-recourse states, we put forth the following second hypothesis:

Hypothesis 2: *The impact difference of house price decline on the mortgage default risk of households is more significant for top-tier homes that for bottom-tier homes.*

2.4 Data and summary statistics

This study uses a sample of 133 metropolitan statistical areas (MSAs) from 32 U.S. states. The states are categorized into recourse and non-recourse ones using the classification of Ghent and Kudlyak (2011) presented in Table 2.1. The locations of all MSAs and states included in our sample are drawn in Figure 2.1. All variables employed in the study are observed monthly from January 2004 to April 2017. The timeframe covers the 2007-2008 financial crisis and the subsequent recovery phase, allowing a comprehensive analysis of the bidirectional relationship between house price change and mortgage default risk over the economic cycle.

Table 2.1: Classification of states according to the state-level recourse laws.

States	Recourse or Non-Recourse	States	Recourse or Non-Recourse
Alabama	Recourse	**North Carolina**	Non-Recourse
Arkansas	Recourse	**Nebraska**	Recourse
Arizona	Non-Recourse	**New Jersey**	Recourse
California	Non-Recourse	**Nevada**	Recourse
Colorado	Recourse	**New York**	Recourse
Connecticut	Recourse	**Ohio**	Recourse
Florida	Recourse	**Oklahoma**	Recourse
Georgia	Recourse	**Oregon**	Non-Recourse
Hawaii	Recourse	**Pennsylvania**	Recourse
Iowa	Non-Recourse	**Rhode Island**	Recourse
Illinois	Recourse	**South Carolina**	Recourse
Indiana	Recourse	**Tennessee**	Recourse
Massachusetts	Recourse	**Texas**	Recourse
Maryland	Recourse	**Virginia**	Recourse
Michigan	Recourse	**Washington**	Non-Recourse
Minnesota	Non-Recourse	**Wisconsin**	Non-Recourse

Notes: The table shows the list of U.S. states in our sample. The states are categorized into recourse and non-recourse states, shown in the two *Recourse or Non-Recourse* columns, according to the classification of Ghent and Kudlyak (2011).

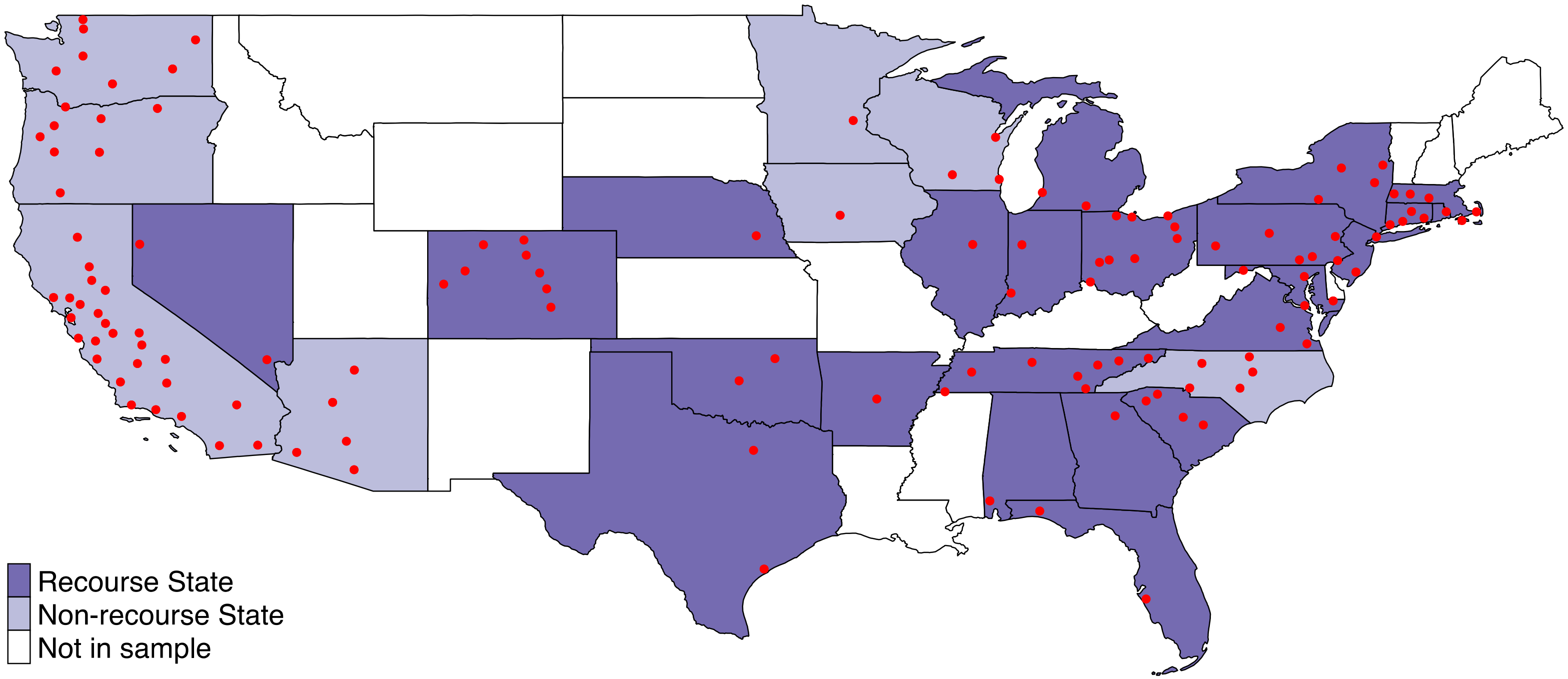


Figure 2.1: Location of metropolitan statistical areas and states.

Notes: The figure presents the locations of the states and metropolitan areas included in our sample. The states are categorized into recourse and non-recourse states according to the classification of Ghent and Kudlyak (2011). The dark purple areas show the locations of recourse states in our sample, while the light purple areas show the locations of non-recourse states. The empty areas show the locations of other U.S. states not included in our sample. The red dots represent the locations of metropolitan statistical areas (MSAs) in the sample. Metropolitan areas in Hawaii are also included in our sample but not shown in the figure.

2.4.1 House price indices and Default risk indicators

We use two house price segments for each metropolitan statistical area: top-tier (TT) and bottom-tier (BT) of the Zillow tiered price indices. These indices capture the median home value for homes that fall into the top and bottom tercile of the house price distribution, respectively, in each MSA. The dynamics of the national-level TT and BT indices are presented in Figure 2.2. On average, the two indices peak in late 2006 and subsequently decline to reach their lowest values in late 2011. Intuitively, top-tier house price reaches its highest and lowest points slightly earlier than bottom-tier price. They recover after that by reaching their pre-crisis period values around 2016. Our analysis uses the log first differences of the two price indices to represent returns.

We use two mortgage default risk indicators: Mortgage Default Risk Index (MDRI) and Homes Foreclosed (HF). MDRI was developed by Chauvet et al. (2016) and is based on Google Search Volume Index (SVI) data for terms such as "foreclosure help" and "government mortgage help" published by Google Trends. We obtained the MDRI index from the UCLA ZIMAN Center for Real Estate.[7] The HF measures the number of homes foreclosed per 10,000 homes each month and is available from Zillow.[8] Specifically, the MDRI reflects the potential default risk of households derived from their online search behaviour, while the HF reflects the actual mortgage default risk of households. The national dynamics of the two mortgage default risk measures are presented in Figure 2.3. Both indicators start to increase in 2007, yet the MDRI shows a sudden surge in the early months of the year, while the HF exhibit a more gradual increase during this year. The MDRI reaches its peak in 2009, while the HF has two local peaks in 2008 and 2011. Further, both measures of mortgage default risk have declined steadily since 2011, falling back in 2016 to their original values registered at the beginning of the sample period in 2004. Considering that the MDRI is reported in levels, while the HF is given in percentage terms, we use the growth rate of the MDRI (ΔMDRI hereafter) and the logarithm values of the HF (LogHF hereafter) in our analysis.

[7] As the city-level MDRI data is only available for 20 cities, we use the state-level MDRI data for all the MSAs in a given state. The monthly state-level MDRI data are available at https://github.com/ChandlerLutz/MDRI_Data.
[8] These measures are obtained from https://www.zillow.com/research/data.

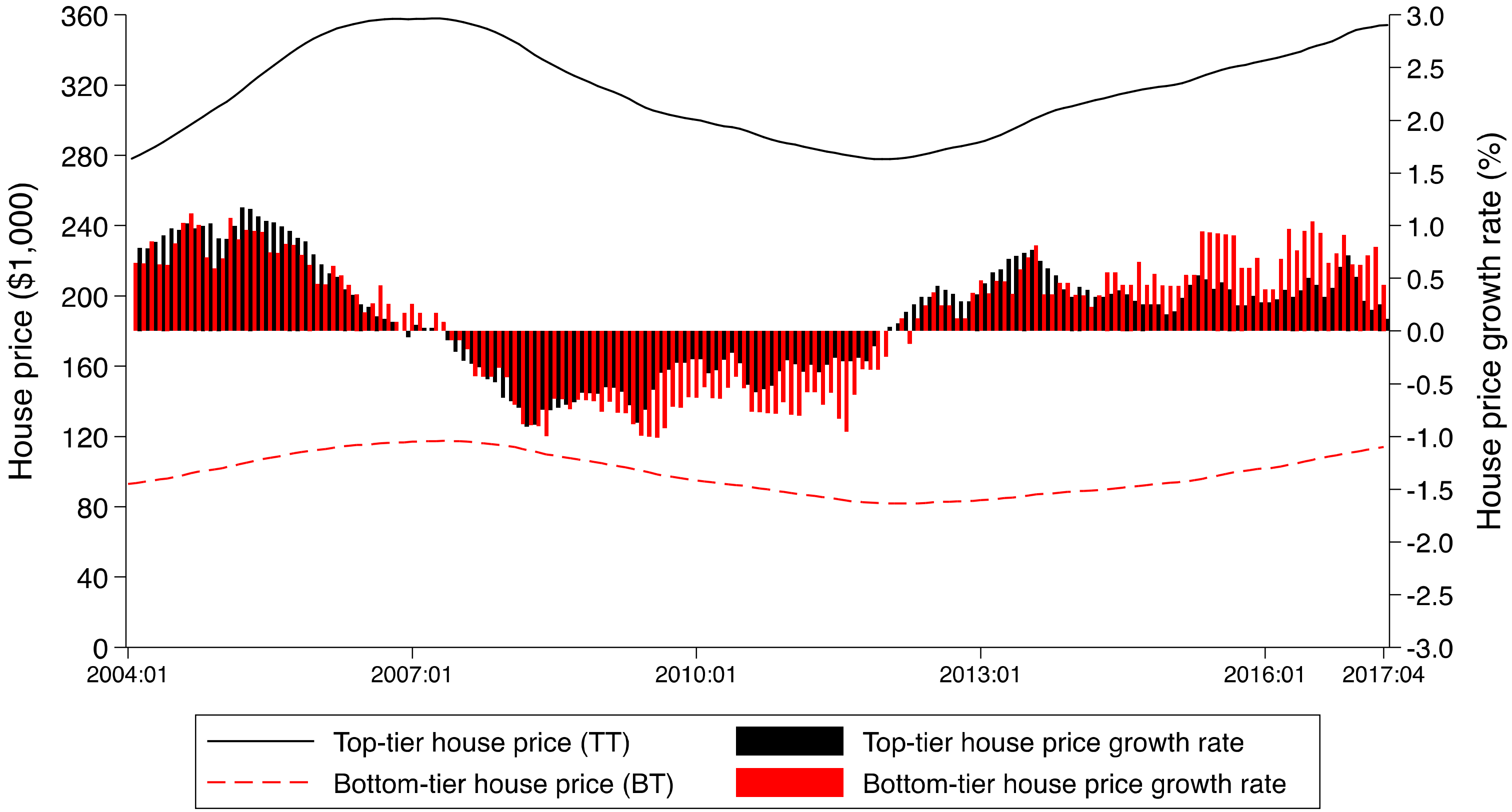


Figure 2.2: Dynamics of national-level top- and bottom-tier house prices and house price returns.

Notes: The figure shows the dynamics of U.S. national-level house prices in the top and bottom-tier housing market segments and the corresponding house price returns of the two indices. The black solid line represents the dynamics of top-tier house price (TT), while the red dotted line represents bottom-tier house price (BT). The left axis shows the range of house prices, and the measurement unit is 1,000 dollars. The black and red bars represent the monthly growth rate of top and bottom-tier house prices, respectively. The right axis shows the range of the growth rate of house prices.

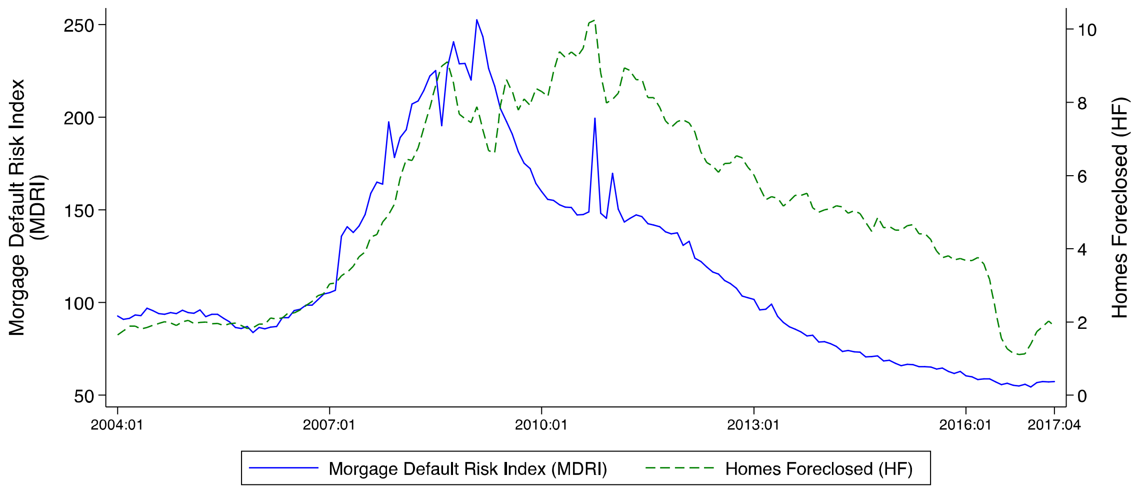


Figure 2.3: Dynamics of national-level mortgage default risk indicators.

Notes: The figure presents the dynamics of two mortgage default risk indicators at the national level of the U.S., in which the solid line represents the dynamics of the Mortgage Default Risk Index (MDRI), and the dashed line represents the dynamics of Homes Foreclosed (HF).

2.4.2 Macroeconomic indicators

To control for the impact of macro-economic factors, we include the total nonfarm employees (EMP), the new private housing units authorized by building permits (PERM), the industrial production index (INDPRO), the producer price index for all commodities (PPIACO), the University of Michigan consumer sentiment (UMCSENT), the effective federal funds rate (FEDFUND) and the S&P 500 Index (SP500) as control variables in our analysis. EMP and PERM are observed at the state level, while INDPRO, PPIACO, UMCSENT, and FEDFUND are available nationally. Except for the SP500, the data for the control variables are provided by the Federal Reserve Bank of St. Louis Economic Data, with all data being collected through Datastream. In our analysis, we use the log difference of all the control variables except for FEDFUNDS, for which we utilize the original level data in our regression models.

Descriptive statistics of all the variables are presented in Table 2.2, and variable definitions are contained in Table A1 in Appendix A.[9]

[9] The descriptive statistics for all variables across subsamples are provided in Table A2 in Appendix A.

Table 2.2: Full-sample descriptive statistics.

Variables	Abbr.	N	Mean	Max	Min	Std. Dev.	ADF test	Transformation	Geographic regions
Top-tier House Price (1,000$)	TT	20865	380.14	1807.80	104.50	251.90	-55.94***	Log first-difference	Metro
Bottom-tier House Price (1,000$)	BT	20527	146.65	657.30	36.90	94.76	-51.55***	Log first-difference	Metro
Mortgage Default Risk Index	MDRI	21280	122.53	627.27	13.74	65.67	-93.59***	Log first-difference	State
Homes foreclosed (%)	HF	17975	6.52	196.08	0.02	8.09	-21.67***	Logarithm	Metro
Employment (1,000)	EMP	21280	5775.30	16697.50	455.90	5043.00	-51.39***	Log first-difference	State
Building Permit	PERM	21280	3580.69	29849.58	28.14	3772.42	-93.71***	Log first-difference	State
Industrial Production Index	INDPRO	21280	97.01	103.60	84.73	4.34	-64.93***	Log first-difference	National
Producer Price Index	PPIACO	21280	183.03	208.30	141.40	18.31	-60.85***	Log first-difference	National
Consumer Sentiment	UMCSENT	21280	81.22	103.80	55.30	11.56	-93.71***	Log first-difference	National
S&P 500 Index	SP500	21280	1472.76	2384.20	735.09	392.01	-90.06***	Log first-difference	National
Federal Funds Rate	FEDFUND	21280	1.35	5.26	0.07	1.81	-14.51***	original value	National

Notes: The table reports the descriptive statistics of all variables in the Panel VAR system. Column *ADF test* gives the value of the Z-statistics from the ADF test for the data after transformation. Specifically, the ADF test is conducted with drift and lag 1 setting. *** denote the null hypothesis that all panels contain unit roots is rejected at 1% statistical levels according to the Z-statistics and p-value from the ADF test. The last column indicates the geographical level at which the variables are observed.

2.5 Methodology

We estimate a Panel Vector Auto Regressive (Panel VAR) model as described by Holtz-Eakin et al. (1988). This approach allows us to focus on the short-run dynamic interaction between house price returns and mortgage defaults, controlling for the macroeconomic environment. We estimate the following system of equations[10]

$$\Delta \boldsymbol{Y}_{i,t} = \begin{pmatrix} \Delta TT_{i,t} \\ \Delta BT_{i,t} \\ DR_{i,t} \end{pmatrix} = \sum_{j=1}^{p} \boldsymbol{A}_j \cdot \Delta \boldsymbol{Y}_{i,t-j} + \boldsymbol{B} \cdot \boldsymbol{X}_{i,t} + \boldsymbol{\varepsilon}_{i,t} \tag{2.1}$$

Whereby p is the order of the Panel VAR model, and $i = 1,2,\dots,N$ denotes the metropolitan statistical area and $t = 1,2,\dots,T$ denotes the month. $\Delta \boldsymbol{Y}_{i,t-j}$ denotes the vector of endogenous variables, which include the appreciation rates of the price indices for two tiers and one of the default risk measures. Here $\Delta TT_{i,t} = \log(TT_{i,t}) - \log(TT_{i,t-1})$ denotes the monthly price appreciation rate (continuously compounded return) of top-tier homes. The price appreciation rate of bottom-tier homes, $\Delta BT_{i,t}$, is defined analogously. The mortgage default risk variable, denoted by $DR_{i,t}$, signifies either the foreclosure sentiment, $\Delta MDRI_{i,t} = \log(MDRI_{i,t}) - \log(MDRI_{i,t-1})$, or the logarithm value of the home foreclosed rate $LogHF_{i,t}$. We estimate model (2.1) separately for the two versions of the mortgage default risk measure. The vector $\boldsymbol{X}_{i,t} = (Emp_{i,t}, Perm_{i,t} Indpro_{i,t}, Umcsent_{i,t}, Ppiaco_{i,t}, Fedfund_{i,t}, SP500_{i,t})$ is the vector of exogenous control variables as previously defined. $\boldsymbol{\varepsilon}_{i,t}$ denotes the vector of idiosyncratic errors. We estimate a Panel VAR system with three lags, which is the optimal number of lags determined by the Schwarz Bayesian Criterion.

In the presence of lagged dependent variables in the model, the commonly used least squares estimator will be biased even when the sample size is large (Judson and Owen, 1999). Therefore, following Arellano and Bover (1995) and Love and Zicchino (2006), we estimate the coefficients $\boldsymbol{A}_j$ and $\boldsymbol{B}$ using the Generalized Method of Moments (GMM), with the lags of the endogenous variables used as instruments. Furthermore, following the method of Abrigo and Love (2016), we apply a Helmert transformation to control for the MSA fixed effects. We report Granger causality tests, impulse response functions, and forecast error variance decompositions for the bidirectional relationship between default risk and housing returns.

[10] A vector autoregressive model with a similar structure, where exogenous variables enter the regression equation contemporaneously, has been analyzed by Yan et al. (2016). The panel specification we use here also considers the cross-sectional dependence across MSAs. We have checked the system stationarity of our model via the AR roots graph.

2.6 Results

In this section, we explore how top-tier and bottom-tier house price depreciation rates respond to shocks to mortgage default risk and, conversely, how mortgage default risk responds to shocks to house price depreciation rates. Our results are based on Granger causality tests, impulse response analysis, and forecast error variance decompositions.

2.6.1 Granger causality tests

Granger causality tests for the interaction between mortgage default risk and tiered house prices are reported in Table 2.3, with the null hypothesis that mortgage default (house price depreciation rate) does not Granger cause house price depreciation rate (mortgage default risk). Panel A tabulates the test results for the null hypothesis that mortgage default risk does not Granger cause house price depreciation rate, while Panel B reports the test results for the null hypothesis that house price depreciation rate does not Granger cause mortgage default risk.

We observe that the top-tier house price depreciation rate Granger causes ΔMDRI in all samples, while the bottom-tier house price depreciation rate only Granger causes ΔMDRI in recourse states. One potential explanation is that high-value homeowners are more active in looking for help online when they observe declines in the value of their homes. Further, the top-tier house price depreciation rate Granger causes LogHF in both the full sample and non-recourse states but not in recourse states. That is, the impact of the top-tier house price depreciation rate on mortgage default depends on foreclosure law in a way consistent with the option theory of default: borrowers in non-recourse states are more likely to walk away from their investments when house prices decline. In comparison, the bottom-tier house price depreciation rate does not Granger cause HF across all samples.

According to the results in Panel B, the null hypothesis that the ΔMDRI does not Granger cause top-tier house price can be rejected for the full sample and in recourse states, but not in non-recourse states. The null hypothesis that the ΔMDRI does not Granger cause the bottom-tier house price depreciation rate cannot be rejected for all samples. Further, LogHF Granger causes the top-tier house price in non-recourse states but not in recourse states. These findings align with our hypothesis that homeowners can strategically choose whether to default after gathering relevant information through Google searches. According to this hypothesis, foreclosures are less likely to occur in recourse states because foreclosure is more costly for borrowers. If high-value homeowners understand these incentives, they will avoid foreclosures in recourse states by choosing other disposition options. This reasoning can explain why, in recourse states, ΔMDRI negatively impacts the prices of high-value homes, but HF does not.

Table 2.3: Granger causality tests.

Causal direction	Full sample			Recourse States			Non-Recourse States		
	Chi^2	p-value	Significance	Chi^2	p-value	Significance	Chi^2	p-value	Significance
Panel A: House price → Mortgage default risk									
ΔTT → ΔMDRI	56.11	0.000	***	9.78	0.021	**	55.16	0.000	***
ΔBT →ΔMDRI	2.30	0.514		12.93	0.005	***	5.59	0.133	
ΔTT → LogHF	19.69	0.000	***	39.33	0.252		39.00	0.000	***
ΔBT → LogHF	6.81	0.078	*	0.00	0.741		5.54	0.136	
Panel B: Mortgage default risk → House price									
ΔMDRI → ΔTT	16.60	0.001	***	13.43	0.004	***	3.90	0.273	
ΔMDRI → ΔBT	5.66	0.130		7.57	0.056	*	1.96	0.581	
LogHF → ΔTT	26.20	0.000	***	4.20	0.240		39.14	0.000	***
LogHF → ΔBT	44.90	0.000	***	51.41	0.000	***	16.43	0.001	***

Notes: The table reports the Granger Causality test results based on regression for the full sample, and two sub-groups (recourse states and non-recourse states). Column *Causal direction* shows the specific Granger causal relationship to be tested. The null hypothesis is that house price (mortgage default risk) does not Granger cause mortgage default risk (house price). Panel A reports the test results for the null hypothesis that house price depreciation rates do not Granger cause mortgage default risk, while Panel B reports the test results for the null hypothesis that mortgage default risk does not Granger cause house price depreciation rates. *, **, and *** respectively indicate that the null hypothesis is rejected at 10%, 5%, and 1% statistical level.

By contrast, default risk measured by the HF Granger causes bottom house price tier both in recourse and non-recourse states, showing no evidence of strategic default behaviour among low-value homeowners.

2.6.2 Impulse response functions

In this section, we quantify the effect of shocks to each endogenous variable on the future dynamics of the other variables through the analysis of impulse response functions. As the traditional impulse response function is dependent on the ordering of endogenous variables within the VAR system, we use the generalized impulse response functions (GIRF) proposed by Pesaran and Shin (1988) instead, which are invariant to the ordering of variables in VAR (see, e.g., Lütkepohl, 2005, p.61, for the criticism of the traditional approach).

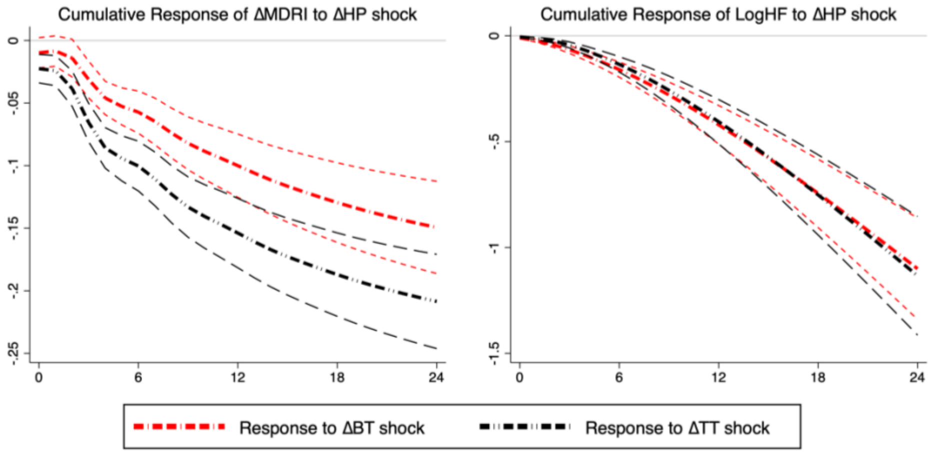
To facilitate comparison to previous studies (e.g., Calomiris et al., 2013), we report standardized impulse functions obtained by dividing the original responses by the sample standard deviation of the corresponding response variables. This also allows for a comparison of the magnitude of the responses across variables.

Figure 2.4 presents the standardized cumulative GIRF of mortgage default risk to a one standard deviation positive[11] shock to the rate of change in the bottom- and top-tier house prices, respectively, as well as the 95% confidence intervals calculated by Monte Carlo simulation. We also present the simple (non-cumulative) impulse responses in Figure A1 in the Appendix. Panel A of Figure 2.4 presents the impulse responses for the full sample, showing that the cumulative responses of ΔMDRI are significantly stronger to a shock to ΔTT than to ΔBT, while the responses of LogHF show no significant differences. Quantitatively, a one-standard-deviation price depreciation rate shock causes a rise in the mortgage default risk, as measured by ΔMDRI, by 14.93% of its sample standard deviation for bottom-tier homes and by 20.85% for top-tier homes over the 24-month forecast horizon. The corresponding response in the mortgage default risk as measured by LogHF is 110.02% of its sample standard deviation for bottom-tier homes and 113.34% for top-tier homes.

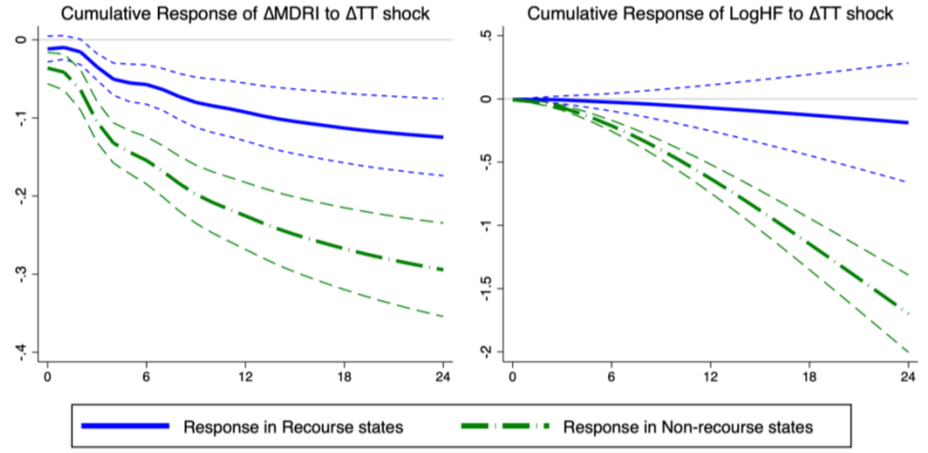
Panels B and C in Figure 2.4 depict the cumulative impulse responses of mortgage default risk measures to house price return shocks in recourse and non-recourse states, respectively, and their 95% confidence intervals. The responses of mortgage default risk to a shock to the house price depreciation rate of top-tier houses are notably stronger in non-recourse states, where default is less costly than in recourse states, for both measures of default risk. A one-

[11] By construction, the effect of a one-standard-deviation negative shock has the same size and the opposite sign.

Panel A. Cumulative response to ΔTT and ΔBT shocks for the full sample



Panel B. Cumulative response to the ΔTT shock in recourse and non-recourse states



Panel C. Cumulative response to the ΔBT shock in recourse and non-recourse states

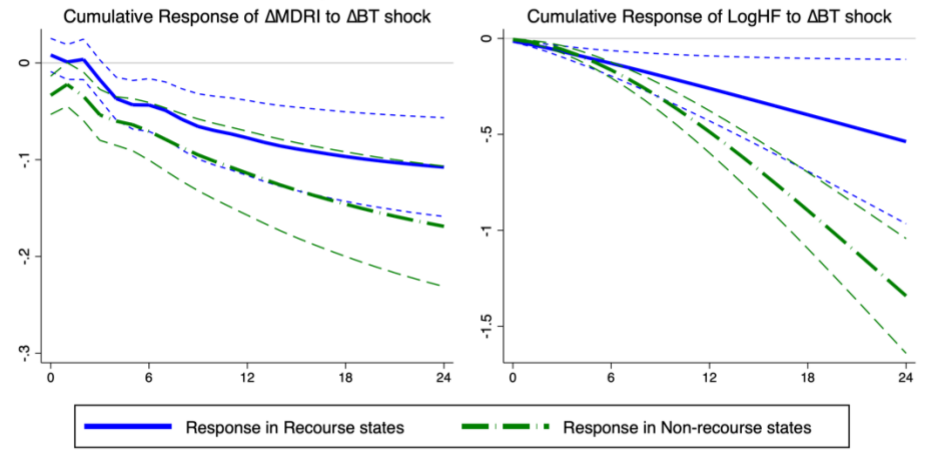


Figure 2.4: Cumulative standardized impulse response of mortgage default risk to shocks to house price returns.

Notes: The thick solid lines represent the cumulative standardized impulse responses of mortgage default risk to shocks to house price returns in the next 24 months. The thin dashed lines represent the 95% confidence interval around the responses. Panel A shows the response of mortgage default risk to shocks to house price returns for the full sample. Panel B shows the response of mortgage default risk to a shock to top-tier house price return (ΔTT), and Panel C shows the response of mortgage default risk to a shock to bottom-tier house price return (ΔBT). The left and right parts of each panel show the responses of mortgage default risk that is measured by the MDRI and HF, respectively.

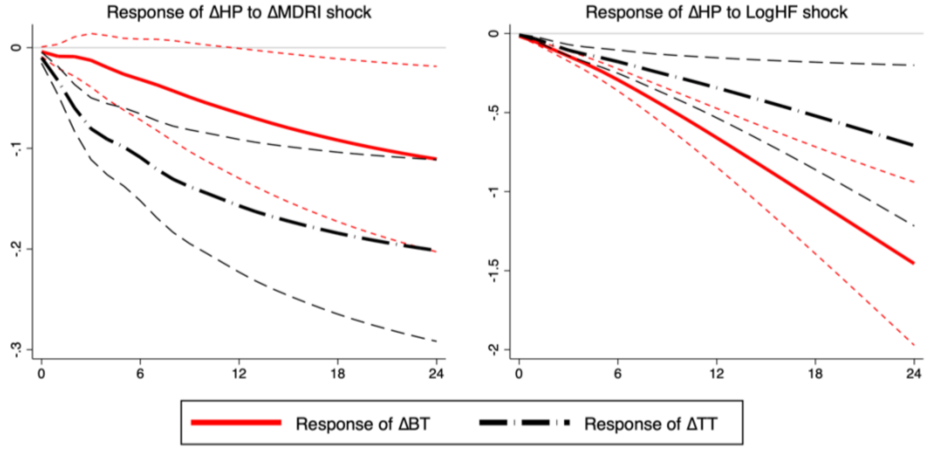
standard-deviation price depreciation rate shock for top-tier homes increases the default risk, as measured by ΔMDRI, by 12.47% of its sample standard deviation in recourse states and by 29.43% of its sample standard deviation in non-recourse states over the next two years, respectively, while the effects on LogHF are 18.81% and 169.80% of its sample standard deviation in recourse and non-recourse states, respectively. In other words, the projected cumulative effect on LogHF in non-recourse states is approximately 9 times greater than in recourse states. The immediate impulse reactions depicted in Panel B of Figure A1 in the Appendix further indicate that mortgage default risk reacts more intensely to shocks from top-tier house price depreciation in non-recourse states.

In comparison, a one-standard-deviation price depreciation rate shock for bottom-tier homes increases the mortgage default risk, as measured by the ΔMDRI, by 10.76% and 16.89% of its sample standard deviation in recourse and non-recourse states, respectively, over the next two years. According to Figure 2.4, the response difference of the ΔMDRI across recourse and non-recourse states is not statistically significant as the response is within the 95% confidence intervals of each other. Furthermore, the corresponding cumulative responses in LogHF are 53.73% and 134.16% of its sample standard deviation in recourse and non-recourse states, respectively. That is, for bottom-tier houses, the effect size is 2.5 times larger in non-recourse states relative to recourse states.

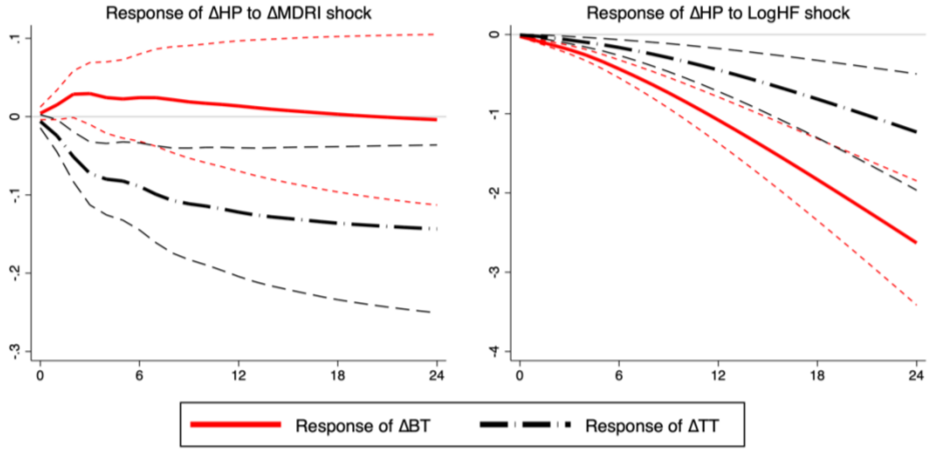
These results indicate that owners of both top- and bottom-tier homes are more likely to default strategically in non-recourse than those in recourse states, while the difference is more significant for top-tier homes. These effects corroborate the results of Ghent and Kudlyak (2011) derived from loan-level data, who report that borrowers are 30% more likely to default in non-recourse states and the deterrent effect of lender recourse is more significant for high value homes.

Figure 2.5 presents the cumulative responses of house price depreciation rate to a one-standard-deviation shock to mortgage default risk and their 95% confidence intervals over the 24-month forecast horizon. Panels A, B, and C depict the results for the full sample, recourse states, and non-recourse states, respectively. We find that top-tier house price returns respond stronger to a ΔMDRI shock than bottom-tier house price returns, while bottom-tier house price returns respond stronger to shocks to LogHF (see Panel A). The above pattern also appears in recourse states (see Panel B). According to Panel B of Figure 2.5, in recourse states, a one-standard-deviation shock of the ΔMDRI leads to a 0.39% and 14.34% of the sample standard deviation change in the bottom- and top-tier house price depreciation rate over a 24-month

Panel A. Cumulative responses of house price returns for the full sample



Panel B. Cumulative responses of house price returns in Recourse States



Panel C. Cumulative responses of house price returns in Non-recourse States

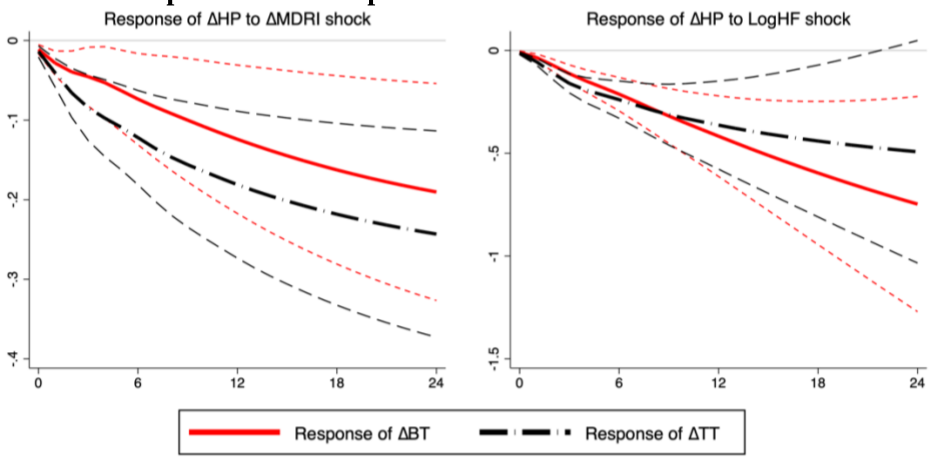


Figure 2.5: Cumulative standardized impulse responses of house price returns to shocks to mortgage default risk.

Notes: The thick solid lines represent the cumulative standardized impulse responses of house price returns to shocks to mortgage default risk in the next 24 months. The thin dashed lines show the 95% confidence interval around the responses. Panels A, B and C show the results for the full sample, recourse states and non-recourse states, respectively. The left and right parts of each panel show responses of house price returns when mortgage default risk is measured by the MDRI and HF, respectively.

forecast horizon, respectively. In comparison, a one-standard-deviation shock of the LogHF leads to a 263.08% and 123.14% of the sample standard deviation change in the bottom- and top-tier house price depreciation rate over a 24-month forecast horizon, respectively.

In non-recourse states, top- and bottom-tier house price returns show no significantly different responses to mortgage default risk measured by either the MDRI or the HF. The response of top-tier (bottom-tier) house price returns to shocks to the ΔMDRI (LogHF) no longer significantly dominates the corresponding response of bottom-tier (top-tier) house price returns. The gap between the responses of top- and bottom-tier house price returns narrows down compared to that in recourse states. According to Panel C, a one-standard-deviation shock of the ΔMDRI (LogHF) results in a decline of ΔBT and ΔTT by 19.03% and 24.32% (74.73% and 49.29%) of its sample standard deviation, respectively, over a 24-month forecasting horizon.

These findings are further supported by the instantaneous impulse responses shown in Figure A2 in Appendix A.

2.6.3 Forecast error variance decompositions

We further use forecast error variance decomposition to analyse the bidirectional relationship between house price returns and mortgage default. Table 2.4 presents the percentage of the forecast error variance of mortgage default risk indices due to innovations in top- and bottom-tier house price returns for 6, 12, 18 and 24 month-ahead forecast horizons. For example, in non-recourse states, at the 24-month-ahead forecast horizon, 0.649% and 20.523% of the forecast error variance of mortgage default risk measured by the ΔMDRI and LogHF can be ascribed to shocks to top-tier house price returns, respectively. In comparison, the corresponding number for recourse states is only 0.123% and 0.092%, respectively.

In the last two columns, denoted by *Non-recourse/Recourse,* we report the relative size of the percentage of forecast error variances explained by innovations in house price returns in non-recourse relative to recourse states. For example, at a 24-month forecast horizon, the percentage of forecast error variance of ΔMDRI and LogHF due to innovations of ΔTT (see Panel A) are 5.3 (0.649%/0.123%) and 223.1 (20.523%/0.092%) times higher in non-recourse states relative to recourse states. These ratios exceed 1 for both ΔMDRI and LogHF and for all forecast horizons. This implies that the top-tier house price returns contribute more to mortgage default risk in non-recourse states than in recourse states. Similarly, according to the results shown in Panel B of Table 2.4, shocks to bottom-tier house price returns also show a greater contribution to mortgage default risk in non-recourse states than in recourse states, although

Table 2.4: Forecast error variance decomposition for mortgage default risk indices.

Horizons	Full sample		Recourse States		Non-recourse States		Non-recourse/Recourse	
(Months)	ΔMDRI	LogHF	ΔMDRI	LogHF	ΔMDRI	LogHF	ΔMDRI	LogHF
Panel A. Percentage of forecast error variance explained by ΔTT								
6	0.211%	0.777%	0.088%	0.024%	0.493%	3.337%	*5.6*	*139.0*
12	0.265%	2.354%	0.111%	0.048%	0.595%	9.742%	*5.4*	*203.0*
18	0.288%	3.999%	0.120%	0.072%	0.633%	15.724%	*5.3*	*218.4*
24	0.298%	5.464%	0.123%	0.092%	0.649%	20.523%	*5.3*	*223.1*
Panel B. Percentage of forecast error variance explained by ΔBT								
6	0.075%	1.028%	0.107%	0.473%	0.204%	1.945%	*1.9*	*4.1*
12	0.108%	2.366%	0.128%	0.561%	0.240%	5.747%	*1.9*	*10.2*
18	0.126%	3.730%	0.136%	0.626%	0.261%	9.584%	*1.9*	*15.3*
24	0.134%	4.947%	0.139%	0.675%	0.272%	12.841%	*2.0*	*19.0*

Notes: The table reports the percentage of the forecast error variance of mortgage default risk indices due to shocks to house prices at the forecast horizons of the 6, 12, 18, and 24 months. Panel A and B report the percentage of forecast error variance explained by top-tier house price and bottom-tier house price, respectively. Column *Non-recourse/Recourse* reports the ratio of the percentage of the forecast error variance for specific mortgage default risk indicator in non-recourse states over the corresponding percentage in recourse states.

the relative size between the explained percentage of forecast error variance in recourse and non-recourse states are much smaller than that due to shocks to top-tier house price returns. Overall, our results speak to the potential of option-based theories of default to explain household behaviour and market reaction to declines in house prices among owners of both low-value and high-value homes.

Table 2.5 presents the percentage of forecast error variance for house price returns due to shocks to mortgage default risk at the 6, 12, 18, and 24-month-ahead forecast horizons. Panels A and B report the impact of shocks to the ΔMDRI and FSR, respectively. The last column (Column *ΔTT/ΔBT*) is the ratio of top-tier house price returns' forecast error variance over bottom-tier house price returns' forecast error variance, owing to shocks to the same mortgage default indicator at different forecast horizons. In all samples, the ΔMDRI shocks explain a greater percentage of the forecast error variance of top-tier house price returns, while LogHF shocks have higher explanatory power for bottom-tier house price returns. Furthermore, in line with the results from impulse response functions, the relative differences are larger in recourse states than in non-recourse states. For instance, in Panel A, the ΔTT/ΔBT ratios are 4.8 and 1.8 for the 24-month forecast horizon in recourse and non-recourse states, respectively. The corresponding ratios in Panel B are 0.3 and 0.5 in recourse and non-recourse states.

2.6.4 An alternative definition of recourse and non-recourse states

The results from previous sections show that borrowers are more likely to default in non-recourse states. According to the hypothesis of this study, the higher mortgage default risk in non-recourse states could be due to the lower default cost in these states due to the lack of recourse by lenders. However, it is also possible that the separation between recourse and non-recourse states is coincidently correlated with some unobserved factors related to the mortgage risk of borrowers. For example, as show in Figure 2.1, most of the non-recourse states in our sample are located on the west coast or in the Midwest of the U.S., with the only exception being North Carolina which is on the east coast. In comparison, most of the recourse states in our sample are located in the Southeast, Northeast, or the eastern areas of the Midwest of the U.S. Therefore, there might be unobserved regional factors driving the impact difference rather than the separation into recourse and non-recourse states.

To test the robustness of our results, we use a finer definition of non-recourse states due to Ghent and Kudlyak (2011). According to their benchmark specification, mortgages are

Table 2.5: Forecast error variance decomposition for house price returns.

Horizons	Full sample			Recourse States			Non-recourse States		
(Months)	ΔTT	ΔBT	ΔTT/ΔBT	ΔTT	ΔBT	ΔTT/ΔBT	ΔTT	ΔBT	ΔTT/ΔBT
Panel A. Percentage of forecast error variance explained by ΔMDRI									
6	0.285%	0.024%	11.9	0.207%	0.046%	4.5	0.372%	0.158%	2.3
12	0.280%	0.042%	6.7	0.198%	0.040%	4.9	0.379%	0.187%	2.0
18	0.276%	0.053%	5.2	0.194%	0.040%	4.8	0.379%	0.202%	1.9
24	0.274%	0.058%	4.7	0.192%	0.040%	4.8	0.379%	0.209%	1.8
Panel B. Percentage of forecast error variance explained by LogHF									
6	0.779%	2.026%	0.4	0.607%	3.640%	0.2	1.595%	1.462%	1.1
12	1.193%	4.493%	0.3	2.294%	11.202%	0.2	1.609%	2.214%	0.7
18	1.729%	7.068%	0.2	5.135%	19.869%	0.3	1.608%	2.741%	0.6
24	2.344%	9.525%	0.3	8.632%	27.970%	0.3	1.608%	3.107%	0.5

Notes: The table reports the percentage of the forecast error variance for tiered house prices due to shocks to mortgage default risk indices at the forecast horizons of the 6, 12, 18, and 24 months. Panel A and B report the percentage of forecast error variance explained by the MDRI and the HF, respectively. In each of the sample groups, the first two columns (i.e., Column *TT* and Column *BT*) indicate the specific house value index of which the variance decomposed, and the last column (i.e., Column *TT/BT*) is the ratio of top-tier house price's forecast error variance over bottom-tier house price's forecast error variance, owing to shocks to the same mortgage default at different forecast horizons.

categorized into recourse and non-recourse based on whether deficiency judgment is explicitly forbidden or impractical in the state. In the new specification, they separate the non-recourse mortgages further into de jure (i.e., explicitly) non-recourse mortgages and de facto (i.e., limited recourse) non-recourse mortgages. As we use state-level mortgage default measures in this study, instead of loan-level data, we separate the non-recourse states in the sample into 'de jure' and 'de facto' non-recourse states in a similar way. Specifically, Arizona, North Carolina, and Oregon are defined as de jure non-recourse states, while California, Iowa, Minnesota, Washington, and Wisconsin are defined as de facto non-recourse states. The locations of states in the new categorization are shown in Figure 2.6. It shows that the locations of de facto or de jure non-recourse states show no regional concentration.

Theoretically, the possibility for lenders to obtain deficiency judgment are highest in recourse states, and lowest in de jure non-recourse states. If the hypothesis that borrowers are less likely to default with consideration of additional default cost resulted from deficiency judgment, we should find the house price depreciation shows strongest impact on the default risk of households in de jure non-recourse states and shows weakest impact in recourse states.

Based on the new state categorization, we compare the impulse responses of mortgage default risk to a one-standard-deviation shock to house price returns in the three state groups. The standardized cumulative GIRF and the 95% confidence intervals calculated by Monte Carlo simulation are presented in Figure 2.7, with Panels A and B representing the responses to a shock to the top- and bottom-tier house price depreciation rate, respectively. In line with theory, the response of mortgage default risk to a house price return shock is strongest in de jure non-recourse states, where deficiency judgment is explicitly forbidden. In contrast, the response is weakest in recourse states, where lenders are more likely to obtain deficiency judgment. In de facto non-recourse states, where deficiency judgment is relatively impractical for lenders to obtain than in recourse states but easier than in de jure non-recourse states, the responses are stronger than in the former but weaker than in the later. Thus, all the results are in line with the hypothesis that borrowers are less likely to default with consideration of additional default cost due to higher possibility of lender recourse. This also suggests that the previous findings regarding the different default risk of households in recourse and non-recourse states is unlikely due to unobserved regional factors.

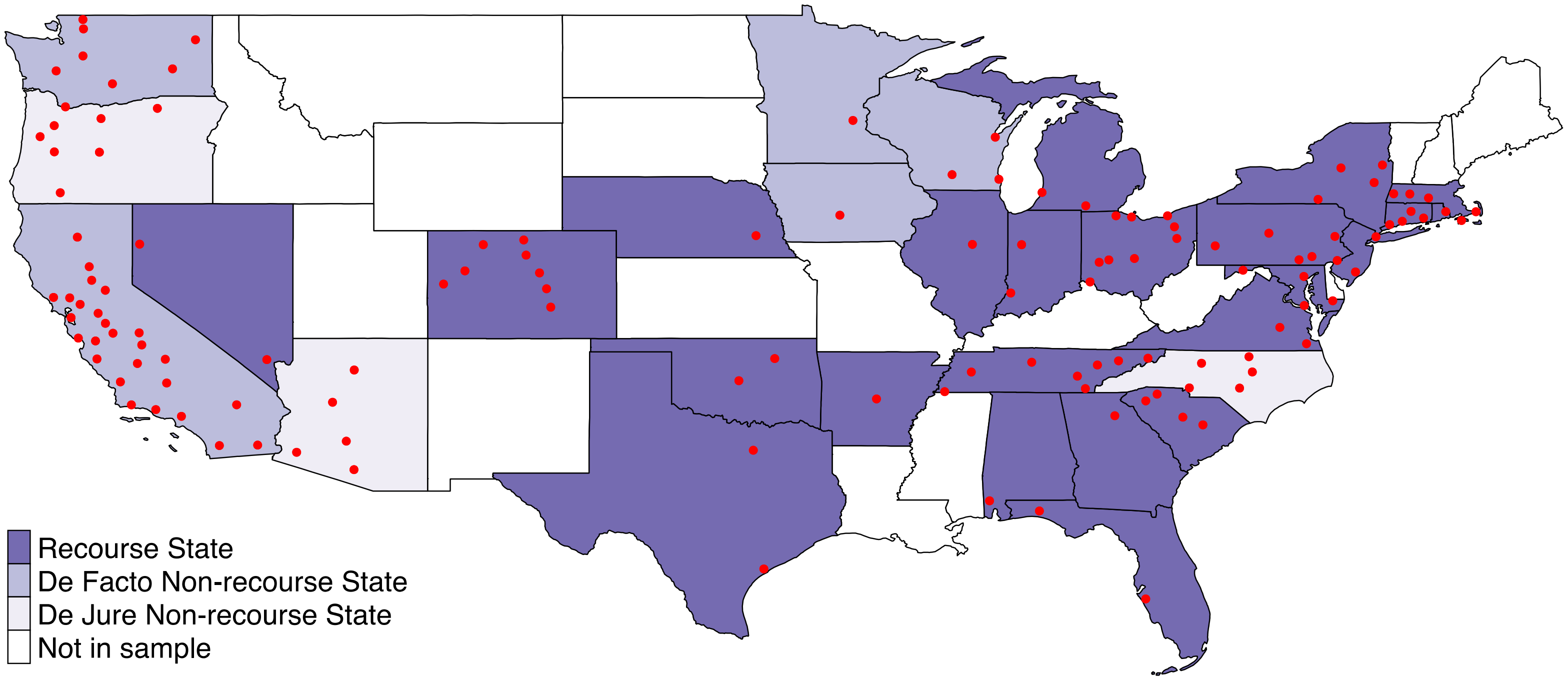
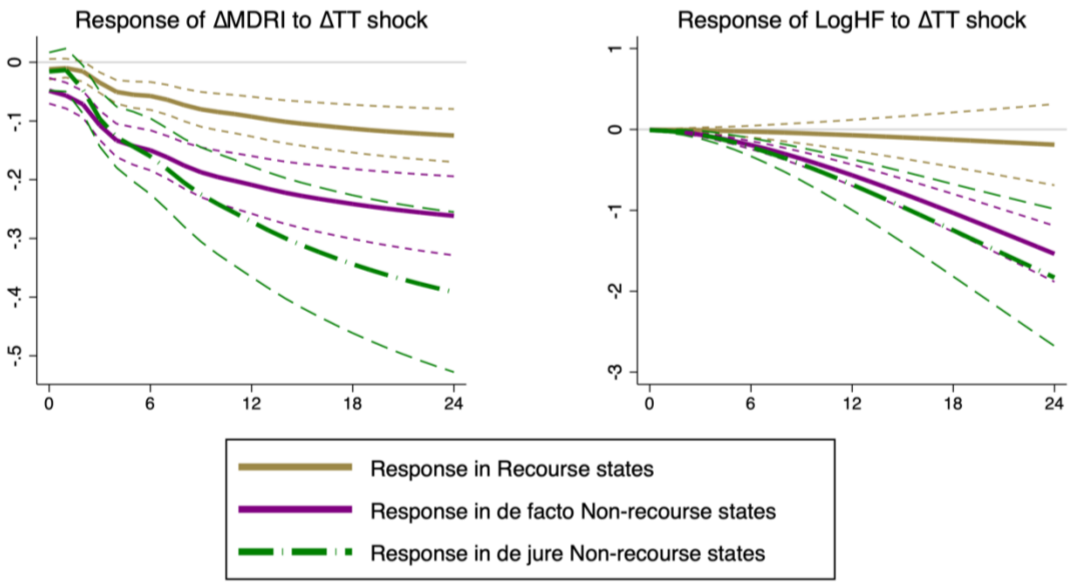


Figure 2.6: Location of metropolitan statistical areas and states in new specification.

Notes: The figure presents the locations of the states and metropolitan areas included in our sample in new specification. The states are categorized into recourse, de facto non-recourse, and de jure non-recourse states according to the classification of Ghent and Kudlyak (2011). The dark purple, middle purple, and light purple areas show the locations of recourse, de facto non-recourse, and de jure non-recourse states, respectively. The empty areas show the locations of other U.S. states not included in our sample. The red dots represent the locations of metropolitan statistical areas (MSAs) in the sample. Metropolitan areas in Hawaii are also included in our sample but not shown in the figure.

Panel A. Cumulative response to the ΔTT shock in recourse, de facto non-recourse, and de jure non-recourse states



Panel B. Cumulative response to the ΔBT shock in recourse, de facto non-recourse, and de jure non-recourse states

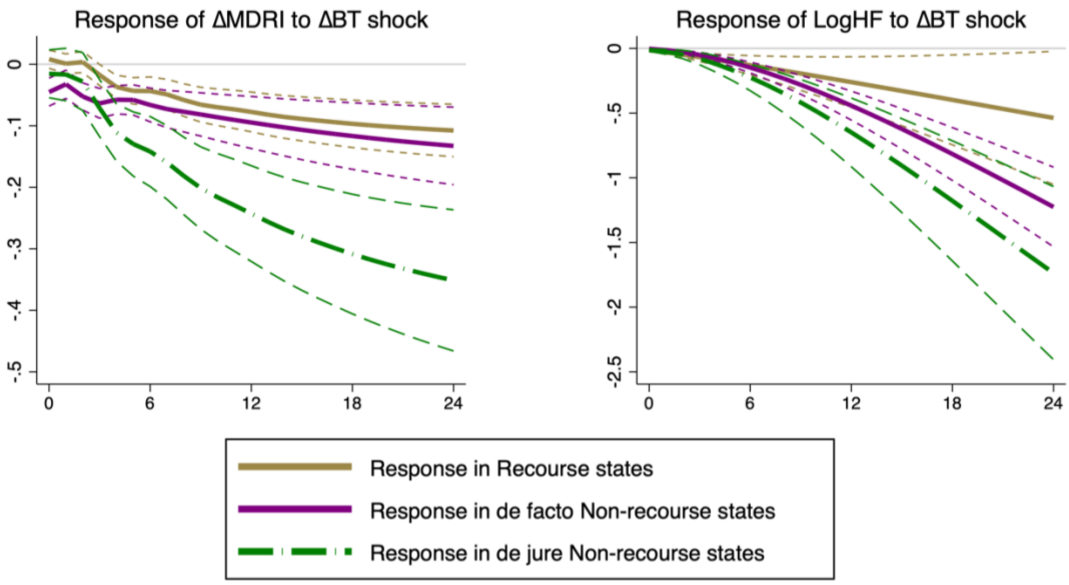


Figure 2.7: Cumulative standardized impulse response of mortgage default risk to shocks to house price returns.

Notes: The thick lines represent the cumulative standardized impulse responses of mortgage default risk to shocks to house price returns in the next 24 months. The thin lines represent the 95% confidence interval around the responses. Panel A shows the response of mortgage default risk to a shock to top-tier house price return (ΔTT), and Panel B shows the response of mortgage default risk to a shock to bottom-tier house price return (ΔBT). The left and right parts of each panel show the responses of mortgage default risk that is measured by the MDRI and HF, respectively.

2.7 Conclusion

This chapter explores the interdependence between mortgage default risk and house price returns in local residential housing markets. We disaggregate regional housing markets into two segments depending on home value (top-tier and bottom-tier) and consider two indicators of mortgage default risk: the Mortgage Default Risk Index (MDRI) based on household sentiment divulged by Google searches and the actual percentage of Homes Foreclosed (HF) in local markets.

We find that house price depreciation rates of both the bottom- and top-tier housing market segments have a significantly stronger impact on mortgage default risk in non-recourse than in recourse states. For instance, a one-standard-deviation price depreciation rate shock for top-tier homes causes ΔMDRI to rise by 12.47% and 29.43% of its sample standard deviation in recourse states and non-recourse states, respectively, and the LogHF to rise by 18.81% and 169.80%, respectively. Similarly, the price depreciation rate shock for bottom-tier homes also causes a greater impact on mortgage default risk in non-recourse states than in recourse states, although the size of the impact difference is smaller than that for top-tier homes. These effects can be interpreted as evidence for the strategic behaviour of homeowners, as default is more sensitive to house price returns when mortgages are non-recourse.

Conversely, the MDRI shows a stronger impact on the price of top-tier homes, while the HF has a stronger impact on the price of bottom-tier homes for the entire sample, as well as in recourse states.

These results are derived in a dynamic macroeconomic equilibrium framework in which the interaction between mortgage default and house prices depends on state- and national-level macroeconomic variables, including employment, industrial production, consumer sentiment, and interest rates. Our results carry implications for mortgage lenders, policymakers, and market regulators in that they suggest that strategic default behaviour is more prevalent at the upper end of the housing market.

Chapter 3

Google search queries, foreclosures, and house prices

3.1 Introduction

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

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

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

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

While this assumption seems plausible, it is unknown how households process the information they gather in their online searches. One possibility, suggested by Chauvet et al. (2016) is that MDRI captures “household concerns about mortgage failure or foreclosure.” An alternative is that households learn by searching for relevant terms online and condition their behaviour on the information they gathered online. That is, as a result of the information they collect online, households may adapt their behaviour when dealing with financial distress, learning how to take advantage of government programs, or interacting with their mortgage lenders. Tetlock (2007) for example, hypothesizes a similar bi-directional relationship when studying the effect of negative media coverage on investor sentiment: While news printed in the Wall Street Journal might convey investor attitudes toward stocks not yet impounded in asset prices, they might also directly shape investors’ perception of stocks.[12] Similarly, online searches might divulge information and at the same time convey information to economic agents who then act on this information. Indeed, top results from online searches include information on government programs to avoid foreclosure as well as legal information. The mechanism of information acquisition by online searchers, however, is different from the one discussed by Tetlock (2007). While the investor reaction to media content described by Tetlock is consistent with noise trader theories implying irrational behaviour, the information gathering by households via online searches could be rational. Online searches can help households chart an optimal plan of action given the legal and institutional framework available in the state in which they reside as well as provide guidance on how to take advantage of government assistance programs.

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

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

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

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

3.2 Literature Review

This paper contributes to two distinct strands of literature. The first strand examines the predictive power of online search intensity on real economic activity. The origin of this research dates back almost a decade when Hal Varian (Google's Chief Economist) suggested that Google Trends data on the search volume for specific keywords helps predict information contained in future government data releases.[13] Since then academics have explored the predictive power of Google's Search Volume Index (SVI) in other domains such as business activity and financial markets. Da et al. (2011, 2015) show that SVI captures investor attention and predicts stock prices at 2-week horizons. Beracha and Wintoki (2013) show that search intensity for terms such as "real estate" and "rent" help predict home prices. Chauvet et al. (2016) construct a mortgage default risk index from the search intensity of SVI for terms such as "mortgage assistance" and "foreclosure help," and show that this index helps predict housing return, mortgage delinquency indicators, and subprime credit default swaps. In this study, we examine the predictive power of this index for city-level housing appreciation rates in different market segments while taking into account local fundamental factors, mortgage market conditions, and mortgage market legislation in the state in which metropolitan areas are located (i.e., whether mortgage contracts are recourse or non-recourse). Specifically, non-recourse states are states in which lenders are not allowed to pursue borrowers for the difference between the mortgage balance owed and the value of their home after homes have been foreclosed. Furthermore, we will also explore the relationship between the MDRI and foreclosure rates, to get a better understanding about identity, reasons, and intentions of the households measured by the index.

The second strand of literature, which developed rapidly in the aftermath of the subprime mortgage crisis, explores the impact of house prices on foreclosure rates. Studies on the contributing role of price declines to mortgage defaults examine the extent to which household behaviour conforms to the "option-theoretical" model of mortgage default. A key prediction of this theory is that households find it optimal to walk away from their investment as soon as their equity falls below a certain (negative) threshold (Foster and Van Order, 1984; Kau et al., 1994). Closely related research on the 'double trigger' hypothesis has developed which aims to disentangle the contributing role of the strategic motive from that of affordability issues and cash flow problems of households (e.g., income shock related to job loss, divorce, or unforeseen healthcare expenses). Empirical studies conducted before the financial crisis find that negative equity is indeed a significant determinant of default (see e.g., Deng et al., 2000; Bajari et al., 2008; and Foote et al., 2008). Using the data from the financial crisis, Elul et al. (2010) present the estimates for the contributions of negative equity, illiquidity (measured by credit card utilization), unemployment shocks and the existence of a second mortgage to the

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

probability of default. More recently, Kelly and McCann (2016) find that short-term arrears are primarily driven by unemployment, negative income shocks or divorce, while long-term arrears are much more likely to be due to negative equity. Using post-crisis data, Mocetti and Viviano (2017) identify job losses as a primary reason for mortgage delinquencies. Ghent and Kudlyak (2011) find that borrowers are 30 percent more likely to default in non-recourse states, whereby this effect is much stronger for homeowners of high-value homes. Moreover, Guiso et al. (2013) use survey data to demonstrate that the willingness to default increases in the home-equity shortfall. Further, the exposure to people who recently defaulted for strategic reasons increases default probabilities because it shows that lenders are unlikely to pursue a deficiency judgment against borrowers. In contrast, Bhutta et al. (2017) find that emotional and behavioural factors are more important in the decision-making process of households than option-theoretic considerations. Gerardi et al. (2018) use data from the Panel Study of Income Dynamics (PSID) to assess the relative importance of negative equity versus the ability to pay. While they find that strategic effects are important, changes in the ability to pay (e.g., job losses) have large estimated effects.

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

The remainder of this paper is organized as follows. In Section 3.3, we present the methodology and in Section 3.4 we describe the data. The empirical results are presented in Section 3.5, and the concluding remarks in Section 3.6.

3.3 Methodology

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

$$P_{i,t}^{j*} = f_j(X_{i,t}) \tag{3.1}$$

where $P_{i,t}^{j*}$ is the logarithm of the fundamental value of the house in tier $j \in \{T, B\}$ (Top and Bottom) in MSA i, in month *t.* Following Abraham and Hendershott (1996) and Capozza et al. (2004), we consider the following components for the vector of fundamental variables $X_{i,t}$: population, income, employment rate, user cost, and construction cost of housing in the MSA. Further, because house prices are also affected by regional geographical and regulatory constraints, we add the land supply elasticity estimates derived by Saiz (2010) as a fundamental factor.[14] These supply elasticity indices vary across MSAs but not across time.

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

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

$$\Delta P_{i,t}^{j} = \alpha_0^j + \alpha_1^j \Delta X_{i,t} + \theta_{i,t}^j \tag{3.2}$$

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

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

$$\theta_{i,t}^{j} = \lambda_0^j + \lambda_1^j \Delta P_{i,t-1}^{j} + \lambda_2^j (P_{i,t-1}^{j*} - P_{i,t-1}^{j}) + \varepsilon_{i,t}^{j} \quad (3.3)$$

In this equation, the coefficient λ_1^j measures the momentum (serial correlation) while the coefficient λ_2^j measures the speed of adjustment to the long-run equilibrium. Combining Equations (3.2) and Equation (3.3) we obtain:

$$\Delta P_{i,t}^{j} = \gamma_0^j + \alpha_1^j \Delta X_{i,t} + \lambda_1^j \Delta P_{i,t-1}^{j} + \lambda_2^j \left(P_{i,t-1}^{j*} - P_{i,t-1}^{j}\right) + \varepsilon_{i,t}^{j} \quad (3.4 - OLS)$$

where $\gamma_0^j = \alpha_0^j + \lambda_0^j$. In addition to an OLS specification, we estimate fixed-effects models that allow for heterogeneity among MSAs and/or time[15]

$$\Delta P_{i,t}^{j} = \gamma_0^j + \alpha_1^j \Delta X'_{i,t} + \lambda_1^j \Delta P_{i,t-1}^{j} + \lambda_2^j \left(P_{i,t-1}^{j*} - P_{i,t-1}^{j}\right) + \vartheta_i + \varepsilon_{i,t}^{j} \quad (3.4 - MSA - FE)$$

$$\Delta P_{i,t}^{j} = \gamma_0^j + \alpha_1^j \Delta X_{i,t} + \lambda_1^j \Delta P_{i,t-1}^{j} + \lambda_2^j \left(P_{i,t-1}^{j*} - P_{i,t-1}^{j}\right) + Year + \varepsilon_{i,t}^{j} \quad (3.4 - Time - FE)$$

$$\Delta P_{i,t}^{j} = \gamma_0^j + \alpha_1^j \Delta X'_{i,t} + \lambda_1^j \Delta P_{i,t-1}^{j} + \lambda_2^j \left(P_{i,t-1}^{j*} - P_{i,t-1}^{j}\right) + \vartheta_i + Year + \varepsilon_{i,t}^{j} \quad (3.4 - MSA\&Time - FE)$$

One difficulty with this estimation is that the fundamental values $P_{i,t-1}^{j*}$ depend on the estimates of the different versions of Equations (3.4) while at the same time they are part of the error correction term which is used as an explanatory variable in these equations. We resolve this issue using the iterative procedure proposed by Abraham and Hendershott (1996). We assume that the observed house price in December 1999 corresponds to its fundamental value (i.e. $P_{i,t}^{j*} = P_{i,t}^{j}$ for t =December 1999) and recover the fundamental value time series from the relationship $P_{i,t}^{j*} = P_{i,t-1}^{j*} + \Delta P_{i,t}^{j*}$. We then re-estimate Equations (3.4) and re-calculate fundamental prices repeatedly until the estimates stabilize (typically we need to perform up to five iterations).[16]

We then analyse how the current (and past) values of the mortgage default risk index, $MDRI_{it}$ impacts the future values of the fundamental component $P_{i,t}^{j*}$ and the bubble component $B_{i,t}^{j} = P_{i,t}^{j} - P_{i,t}^{j*}$ of local house prices as well as the foreclosure rates $HF_{i,t}$.

[15] As the housing supply elasticity is time-invariant for each metropolitan area, in the MSA-level fixed effect regression specifications (4.MSA-FE) and (4.MSA&Time-FE) we exclude this variable from the vector of fundatament factors $\Delta X_{i,t}$ (and denote the resultant vector by $\Delta X'_{i,t}$).

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

Furthermore, we use the house price decomposition to explore how changes in the fundamental $P_{i,t}^{j*}$ and the bubble $B_{i,t}^{j}$ components of home values affect foreclosure rates $HF_{i,t}$.

3.4 Data, variable construction, and summary statistics

The estimation of the fundamental house price model is based on a panel of 107 MSAs located in 29 U.S. states. A map with the location of these MSAs is presented in Figure 3.1. For each MSA we observe the monthly growth rate of house prices and local fundamental factors. Further, in our analysis of the effect of the mortgage default risk index on house prices and foreclosures, we include additional variables that account for the mortgage market conditions in each MSA.

All MSAs in the dataset are listed in Table B1 in Appendix B along with the state in which they are located. The table also classifies the states into recourse and non-recourse category depending on whether states allow lenders to pursue a deficiency judgment against foreclosed borrowers (we use the classification of Ghent and Kudlyak, 2011). The geographical location of the studied MSAs is presented in Figure 3.1. In this figure, the recourse states are depicted in dark blue and the non-recourse states are represented in light blue colour.

3.4.1 Local house prices and fundamental factors

In this study, we use the monthly Zillow home value indices[17] for the period from April 1996 to December 2016. These indices are constructed from deed records using a hedonic methodology which accounts for individual attributes such as the size and the number of bedrooms and bathrooms. A major challenge in the construction of home value indices is the changing composition of the properties sold in different periods. Indices based on a repeat-sales methodology – such as the S&P Case-Shiller index or the index of the Federal Housing Finance Agency – account for this issue by using only properties that are sold more than once. This methodology has limitations for smaller regions or smaller market segments where the number of repeat sales is limited.[18] Zillow, on the other hand, aggregates all transactions to create valuations for all properties (Zestimates) based on their characteristics[19] and uses the Zestimates to construct its regional price indices (see, e.g., Dorsey et al., 2010 for a detailed discussion of this approach).

[17] These data are obtained from https://www.zillow.com/research/data
[18] Indeed, the S&P Case-Shiller index covers only 20 cities.
[19] For more information on the Zillow methodology see https://www.zillow.com/research/zhvi-methodology/

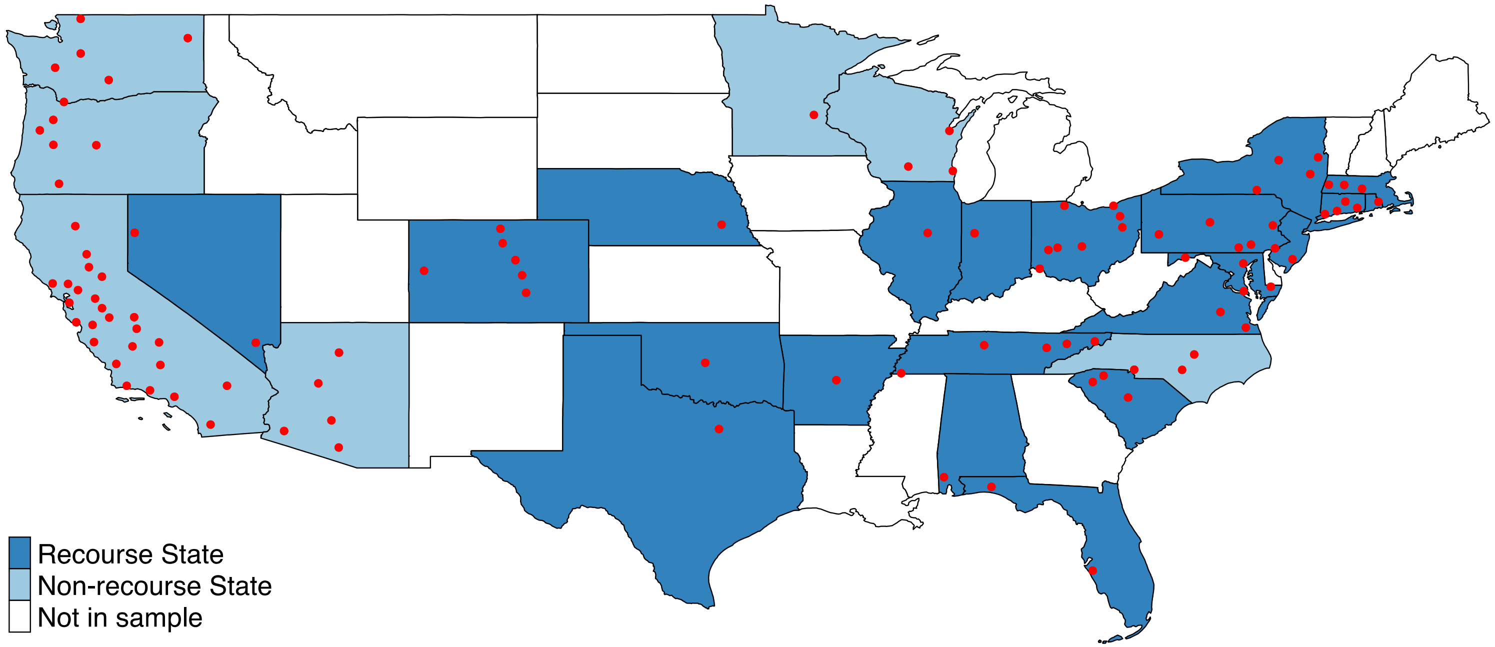


Figure 3.1: Location of metropolitan statistical areas and states.

Notes: The figure presents the locations of the states and metropolitan areas included in our sample. The states are categorized into recourse and non-recourse states according to the classification of Ghent and Kudlyak (2011). The dark blue areas show the locations of recourse states in our sample, while the light blue areas show the locations of non-recourse states in our sample. The empty areas show the locations of other U.S. states that are not included in our sample. The red dots represent the locations of metropolitan statistical areas (MSAs) in the sample. Metropolitan areas in Hawaii are also included in our sample, but not shown in the figure.

As we are interested in the dynamics of different market segments, we use the top and the bottom house price tiers in our analysis. The top tier index captures the median value of homes within the 65^{th} to 95^{th} percentile range while the bottom tier index captures the median value of homes within the 5^{th} to 35^{th} percentile range for each MSA. The dynamics of the top and the bottom price tiers for three of the MSAs in the dataset (San Diego, Minneapolis, and Phoenix) are presented in Figure 3.2 (Panels A and B). These three MSAs represent a cyclical market, steady market, and bubble market, respectively, according to the classification in Mayer's (2011) survey article on housing bubbles.

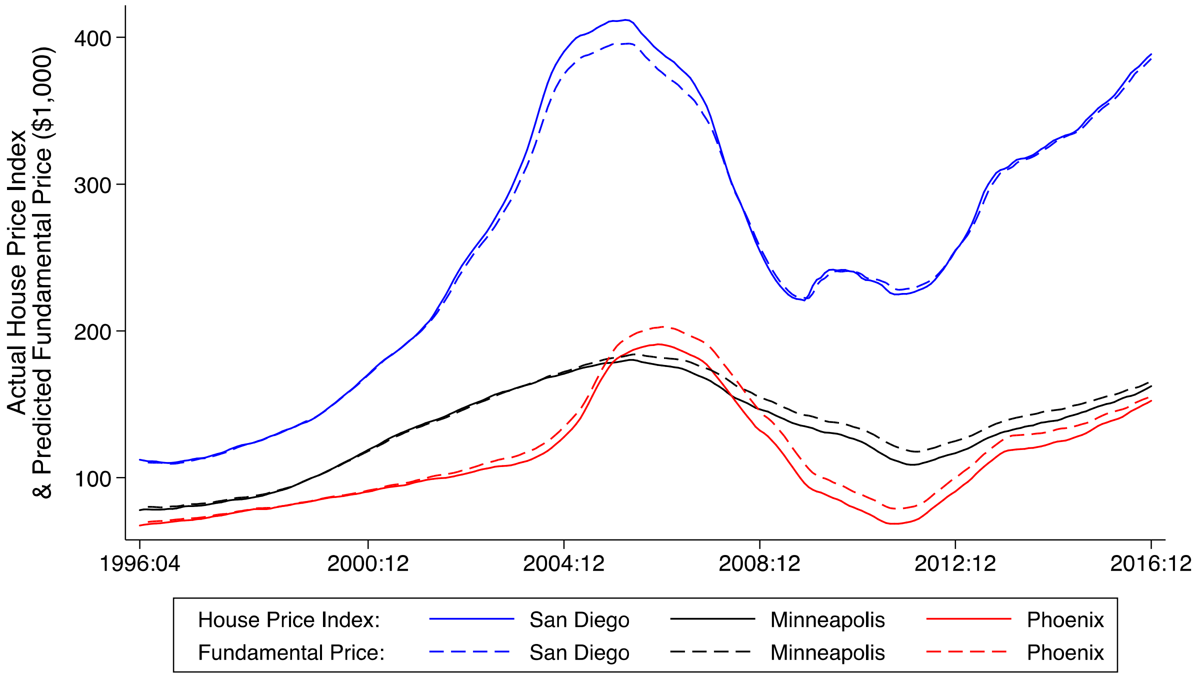
Although there is a variation across regional housing market segments, on average the indices peak in late 2006, and then decline and reach their lowest values between 2009 and 2012. They recover thereafter by almost reaching their pre-crisis period values around 2016. In our analysis, we use the log differences of the price indices (i.e., the continuously compounded returns) for the two market segments.

The fundamental variables used include the population, personal income per capita, total non-farm employment, construction cost, a derived user cost of homeownership, and the land supply elasticity index in the MSA. Descriptive statistics of these variables, except for land supply elasticity which is time-invariant, along with unit root tests are presented in Table 3.1. The population and personal income data are collected from the Bureau of Economic Analysis. We use cubic spline interpolation (Boor, 1978) to derive monthly values from the original annual observations. The total non-farm employment, available at the state level, is collected from Datastream and used for all metropolitan areas located in the same state. The construction cost is measured by the price index of new single-family houses under construction, which is available from the U.S. Census Bureau. As only the national index is available in monthly frequency, the change in construction costs varies over time but not across MSAs. As a measure of land supply elasticity of MSAs, we use the land supply estimates derived by Saiz (2010).[20]

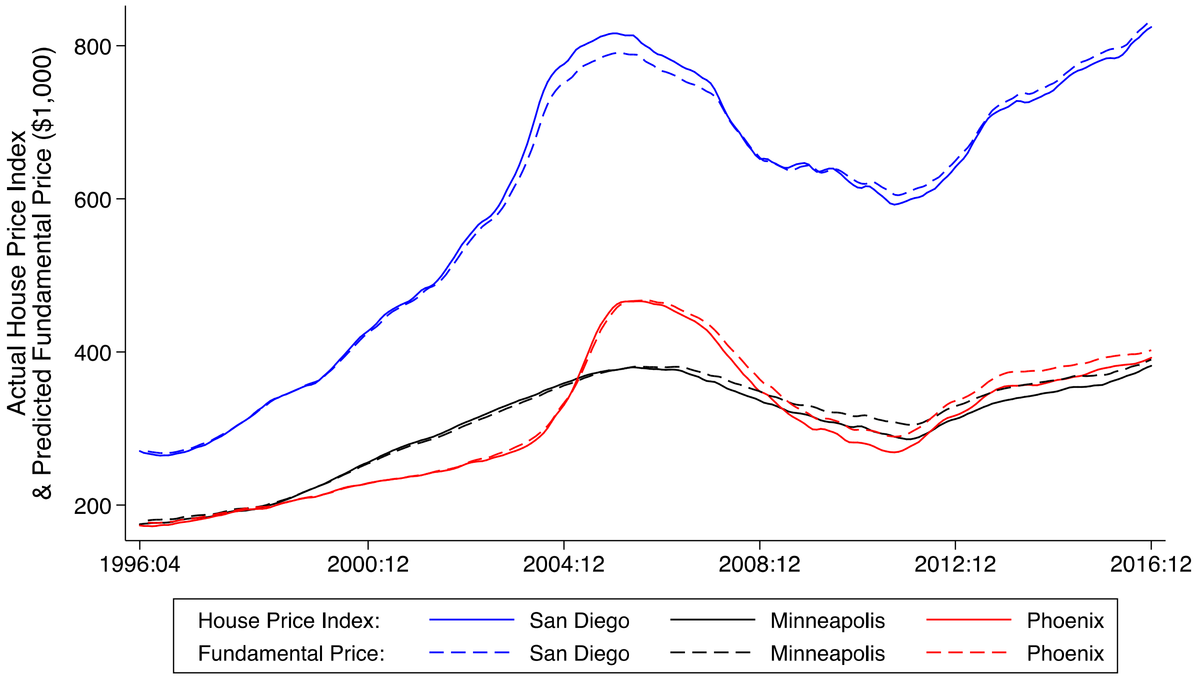
To facilitate comparison to previous research, we construct the user cost by the method of Capozza et al. (2004) which accounts for mortgage rates, taxes, expected appreciation as well as annual maintenance and depreciation of properties. That is, the user cost is constructed by the following formula:

[20] As this elasticity measure has only limited coverage, we are left with only 93 MSAs in our sample. Another way to account for differences across MSAs is to estimate a model with MSA-level fixed effects while leaving out the supply elasticity as a regressor. In Table 3.2 we report results for both the OLS and the fixed effect model, but use the estimates of the fixed effect model in the subsequent analysis because this model allows us to use all 107 MSAs in our sample.

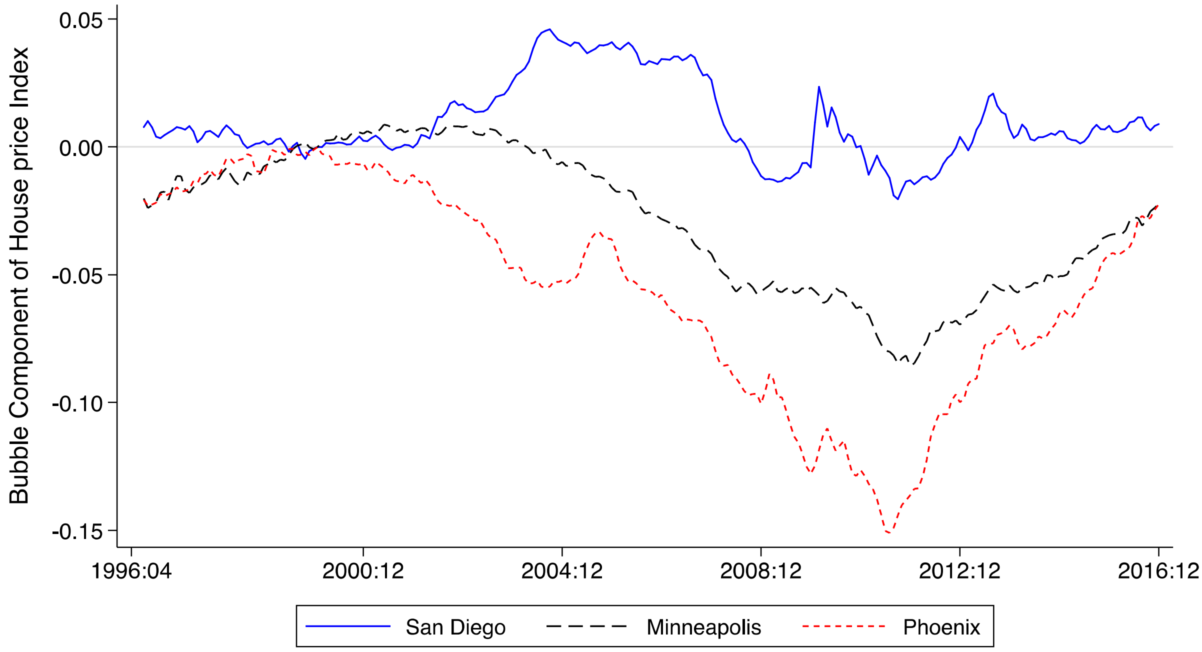
Panel A. Bottom-tier house price index (solid lines) and fundamental price (dash lines)



Panel B. Top-tier house price index (solid lines) and fundamental price (dash lines)



Panel C. Bubble component of the bottom-tier house price index



Panel D. Bubble component of the top-tier house price index

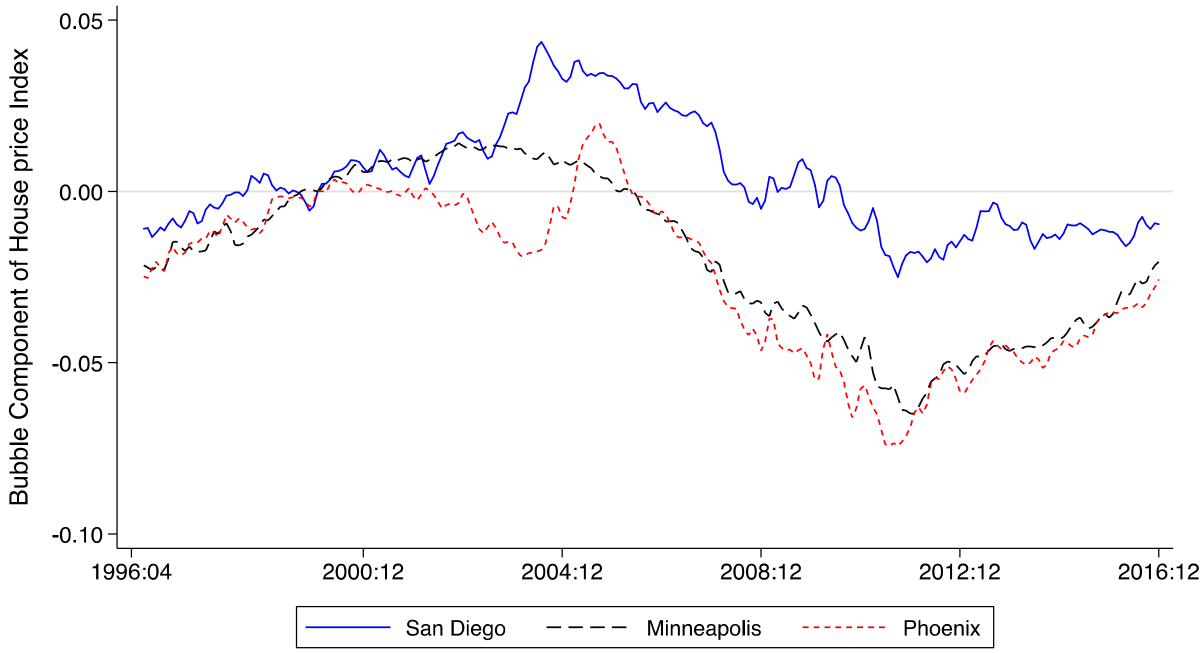


Figure 3.2: Dynamics of house price indices for selected MSAs.

Notes: San Diego, Minneapolis, and Phoenix represent examples of a cyclical, a steady, and a bubble market, respectively, according to the classification of Mayer (2011). The bubble component is calculated as the deviation from the fundamental house price, i.e. the difference between the logs of the house price index and its fundamental component: $B_{i,t}^{j} = P_{i,t}^{j} - P_{i,t}^{j*}$.

Table 3.1: Descriptive statistics.

Variables	Sample period	Mean	Max	Min	SD	ADF test		Transformation
						Level	Transformed	
Top tier house price ($)	Apr 1996 - Dec 2016	320162	1785000	70400	199068	0.88	-48.37***	Log first-difference
Bottom tier house price ($)	Apr 1996 - Dec 2016	127811	636600	28200	83657	-0.87	-45.02***	Log first-difference
Mortgage Default Risk Index	Jan 2004 - Dec 2016	122.13	554.60	13.74	62.98	6.94	-83.8***	Log first-difference
Homes foreclosed (%)	Jan 1998 - Dec 2016	5.46	106.20	0.01	7.35	-9.46***	-4.33***	Logarithm
Employment (1,000)	Apr 1996 - Dec 2016	6097	16638	437	5139	6.5	-51.76***	Log first-difference
Construction cost index	Apr 1996 - Dec 2016	93.43	124.00	67.70	15.03	19.48	-75.74***	Log first-difference
User cost (%)	Apr 1996 - Dec 2016	0.14	0.22	0.10	0.02	-35.82***	-84.05***	First difference
Population (1,000)	Apr 1996 - Dec 2016	1050	13340	77	1607	6.3	-4.6***	Log first-difference
Per Captia Personal income ($)	Apr 1996 - Dec 2016	35932	107936	16425	10536	22.22	-46.15***	Log first-difference
Total loan supply (1,000 $)	Jan 2007 - Dec 2016	7157	118789	155	12067	-1.36*	-6.71***	Log first-difference
Percentage of subprime mortgage (%)	Jan 2007 - Dec 2016	2.59	23.56	0.03	4.58	-11.74***	-6.32***	Logarithm

Notes: The table reports descriptive statistics and stationarity tests. The Top and Bottom house price tiers are measured in U.S. Dollars. The MDRI is constructed from 11 Google Trends search items such as “foreclosure,” “mortgage help,” or “government assistance,” (see Chauvet et al., 2016, Table 1, for the full list). The Homes foreclosed are the number of foreclosures per 10,000 homes. The monthly observations of Population and Personal income are derived from annual data via cubic spline interpolation (Boor, 1978). The User cost is constructed from Equation (3.5). The tests for stationarity are Fisher-type Augmented Dickey-Fuller (ADF) tests for unbalanced panels that have as the null hypothesis that all panels contain a unit root (Choi, 2001). ADF unit root test is conducted on the level value and the transformed value of all variables, with the null hypothesis that all panels contain unit roots. The corresponding transformation methods are given in the last column. *, **, and *** represent the null hypothesis are rejected at the 10, 5, and 1% statistical level, respectively.

$$User\ cost = (Mortgage\ Rate + Property\ Tax\ Rate) \times\ (1 - Income\ Tax\ Rate)$$
$$-Inflation + 0.03 \qquad (3.5)$$

Here the "Mortgage Rate" is the 30-Year fixed-rate mortgage average in the United States, collected from the Federal Reserve Bank of St. Louis. The "Property Tax Rate," collected from Wallethub,[21] is the effective real-estate state tax rate. The "Income tax rate" is the sum of the average federal income tax rate and average state income tax rate for the middle quintile of households. The federal income tax rate is collected from the Urban-Brookings Tax Policy Center,[22] while the state income tax rate is collected from the National Bureau of Economic Research.[23] For inflation, we use the CPI provided by the Federal Reserve Bank of St. Louis. The annual maintenance and obsolescence of properties are set at 3 percent as indicated in Equation (3.5).

3.4.2 Mortgage lending

To account for local mortgage market conditions, we construct two variables from Home Mortgage Disclosure Act (HMDA) data:[24] the total amount of mortgage loans in a given year in each MSA (Loan supply) and the percentage of loans that are subprime, or higher-priced mortgage loans in each MSA (Subprime). Loans are categorized as subprime following the classification of Mayer and Pence (2008) according to which a mortgage is a subprime mortgage if its rate spread exceeds 3 percent for first-lien mortgages and 5 percent for junior lien mortgages.[25]

3.4.3 Mortgage default risk

The Mortgage Default Risk Index (MDRI hereafter) of Chauvet et al. (2016) is constructed from the Google Search Volume Index (SVI) data for terms such as "foreclosure help" and "government mortgage help" in U.S. states published by Google Trends.[26] The MDRI is

---

[21] Property tax rates are collected from: https://wallethub.com/edu/states-with-the-highest-and-lowest-property-taxes/11585/

[22] The average federal income tax rate is downloaded from: https://www.taxpolicycenter.org/statistics/historical-average-federal-tax-rates-all-households

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

[24] The HMDA data contains over 80 percent of home loans and is the most comprehensive source of data on mortgage loans (Avery et al., 2007).

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

[26] For the construction of their monthly MDRI, Chauvet et al. (2016) used "foreclosure assistance+foreclosure help+government assistance mortgage+home mortgage assistance+home mortgage help+housing

obtained from the UCLA ZIMAN Center for Real Estate.[27] Zillow also publishes a Homes Foreclosed index (HF hereafter) which gives the number of homes foreclosed per 10,000 homes in metropolitan areas each month. As an illustration, in Figure 3.3 (see Panels A and B) we present the dynamics of the MDRI and HF in three of the MSAs in our sample – San Diego, Minneapolis, and Phoenix. Both indicators start to increase in early 2007 and reach their peak around 2008 in San Diego, and around 2009 in Minneapolis and Phoenix. The descriptive statistics of these variables are presented in Table 3.1.

3.5 Results

We first explore how real house prices respond to changes in local fundamental factors by estimating the models given in Equation (3.4). In particular, we consider population, personal income, employment, as well as the variable we created for the user cost, construction cost, and the land supply elasticity of the MSA (cf. Capozza et al., 2004; Stevenson, 2008). As a preliminary step, we perform unit root tests on the tiered house price indices as well as the fundamental variables (see the results in the Column *ADF test* in Table 3.1). Most of the variables are non-stationary in levels and stationary after corresponding transformation. This points to the inherent difficulties that would be present if we tried to directly use the levels of these variables in our statistical analysis. Furthermore, it justifies our focus on growth rates and the use of an error correction modelling approach.

3.5.1. Long-run equilibrium relationship

The regression results of the error correction models specified in the four versions of Equation (3.4) are presented in Table 3.2. They include OLS estimates as well as estimates of fixed-effect models in which we control for MSA and time fixed effects.

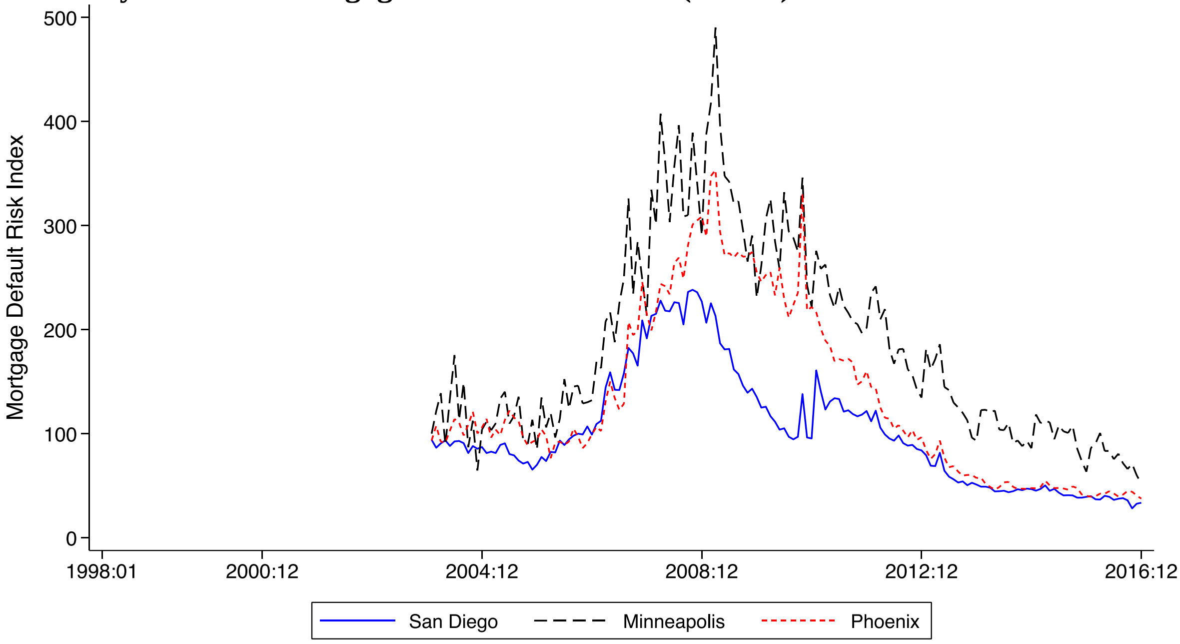
The coefficient estimates for all fundamental variables have the anticipated sign and are statistically significant at the 1 percent or 5 percent level. As expected, growth in population, personal income, and employment has a positive impact on house prices. An increase in user cost, a significant component of which constitutes the mortgage interest rate, is associated with lower house price growth. Similarly, an increase in construction cost leads to an increase in home values. Further, the relationship between the land supply index and house prices is negative, as had been found in previous literature. The error correction term is significant

---

assistance+mortgage assistance program+mortgage assistance+mortgage foreclosure help+mortgage foreclosure+mortgage help" to obtain the joint SVI.

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

Panel A. Dynamics of Mortgage Default Risk Index (MDRI)



Panel B. Dynamics of Homes Foreclosed (HF)

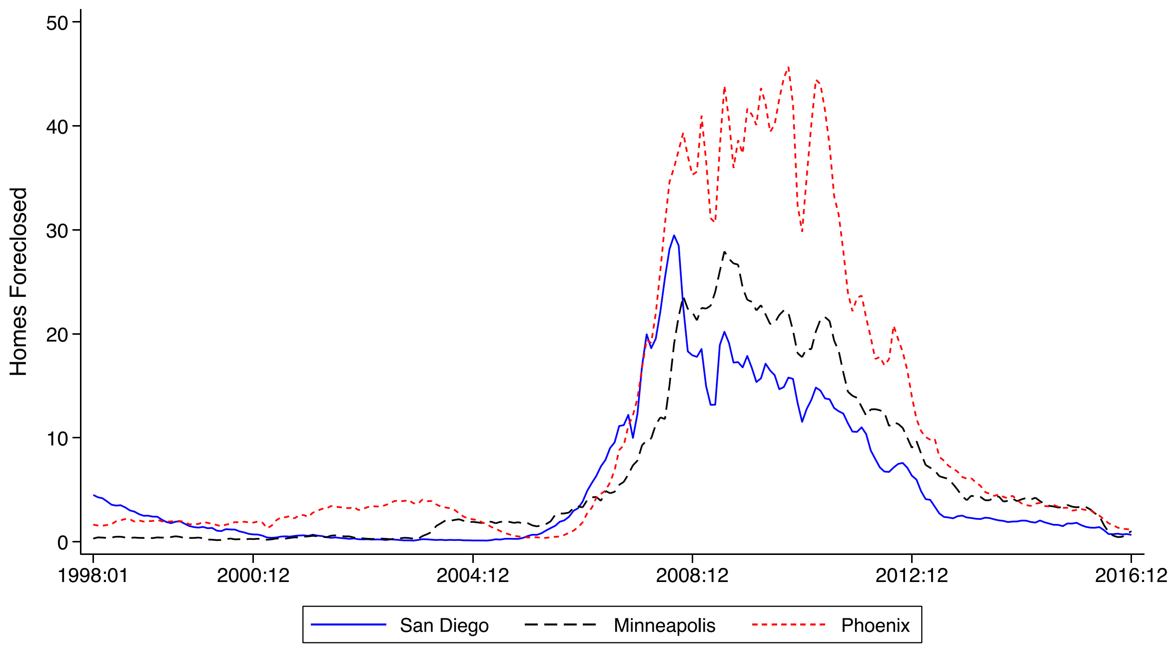


Figure 3.3: Dynamics of default risk indices for selected MSAs.

Notes: Panel A and Panel B represent the dynamics of the Mortgage Default Risk Index (MDRI) and the number of Homes Foreclosed (HF) per 10,000 homes in San Diego, Minneapolis, and Phoenix.

Table 3.2: Estimates of the fundamental house price model.

	Model 1: Pooled OLS		Model 2: MSA Fixed effects		Model 3: Time fixed effects		Model 4: MSA and Time Fixed effects	
	ΔBottom	ΔTop	ΔBottom	ΔTop	ΔBottom	ΔTop	ΔBottom	ΔTop
ΔPopulation	0.0080***	0.0072***	0.0324***	0.0347***	0.0121***	0.0115***	0.0452***	0.0446***
	(2.92)	(3.17)	(8.57)	(10.90)	(4.22)	(4.95)	(11.21)	(13.13)
ΔPersonal income	0.0409***	0.0520***	0.0260**	0.0352***	0.1096***	0.0906***	0.0715***	0.0604***
	(3.57)	(5.48)	(2.42)	(3.93)	(8.21)	(8.21)	(5.81)	(5.85)
ΔEmployment	0.0419***	0.0366***	0.0822***	0.0839***	0.0528***	0.0327***	0.0545***	0.0399***
	(3.70)	(3.91)	(7.76)	(9.50)	(4.12)	(3.10)	(4.55)	(3.98)
ΔHouse Price$_{t-1}$	0.9032***	0.8909***	0.8944***	0.8779***	0.8664***	0.8431***	0.8506***	0.8343***
	(305.82)	(286.93)	(327.16)	(300.84)	(252.91)	(231.94)	(267.82)	(241.12)
Change in user cost	-0.0299***	-0.0228***	-0.0239***	-0.0193***	-0.0130**	-0.0121**	-0.0130**	-0.0137***
	(-5.44)	(-5.02)	(-4.59)	(-4.46)	(-2.25)	(-2.55)	(-2.40)	(-3.02)
ΔConstruction cost	0.0581***	0.0336***	0.0526***	0.0312***				
	(13.43)	(9.40)	(12.77)	(9.10)				
Supply elasticity	-0.0001**	-0.0001***			-0.0001***	-0.0001***		
	(-2.50)	(-2.60)			(-3.25)	(-3.86)		
Error correction$_{t-1}$	0.0073***	0.0070***	0.0319***	0.0369***	0.0056***	0.0050***	0.0369***	0.0277***
	(9.75)	(9.82)	(27.31)	(30.55)	(7.98)	(7.51)	(30.76)	(24.05)
Constant	0.0001	0.0002***	-0.0003***	-0.0002***	-0.0005***	0.0002	-0.0011***	-0.0003**
	(1.24)	(2.96)	(-5.44)	(-5.09)	(-2.78)	(1.58)	(-6.45)	(-2.26)
Number of Obs	22,815	22,804	26,259	26,248	22,815	22,804	26,259	26,248
Adjusted R-squared	0.8425	0.8242	0.8399	0.8205	0.8462	0.8292	0.8452	0.8240

Notes: The table presents regression results of the fundamental house price model defined by the four versions of Equation (3.4). The continuously compounded returns of the Bottom and Top tier house price indices are denoted by ΔBottom and ΔTop, respectively. For the population, employment, personal income, and construction cost variables, the continuously compounded growth rates are used as regressors. Following Abraham and Hendershhott (1996), the change in the user cost is used as a regressor. One, two, and three asterisks represent significance at the 10, 5, and 1%, respectively. Corresponding t-statistics of coefficients are presented in parentheses.

indicating that both the top and the bottom house price tiers adjust to their long-run equilibrium values. Similarly, the autoregressive coefficient is positive and statistically significant, indicating the presence of momentum in housing returns for both house price tiers.

The magnitude of the coefficients suggests that bottom tier homes are more sensitive to changes in population, employment, user cost, and construction cost as well as exhibit a stronger momentum. We formally test whether the coefficients for the top tier and the bottom tier are significantly different from each other using the OLS model specification (Model 1 in Table 3.2). In particular, we construct a dummy variable "Toptier," which takes on the value of one for the top tier and zero for the bottom tier index. We include it as a regressor along with the interactions of this variable with the fundamental variables. We estimate this regression using Abraham and Hendershott's (1996) iterative method described in the methodology section by pooling the top tier and bottom tier observations together. We find that only the coefficients for the interaction variables $(\text{Toptier} * \Delta\text{House Price}_{t-1})$ and $(\text{Toptier} * \Delta\text{Construction cost})$ are significant. They have a negative sign indicating starter homes exhibit a stronger momentum effect and their response to construction cost is greater compared to trade-up homes.

The fundamental house price model allows us to disaggregate house price indices into their fundamental and bubble components. Using the estimates of Model 2 (Panel Fixed Effects) in Table 3.2, we calculate these two components of house price and represent their dynamics for three of the MSAs (San Diego, Minneapolis, and Phoenix) in Figure 3.2. In the following subsections, we analyse whether the MDRI helps predict these components of house prices and whether these components affect future foreclosures.

3.5.2 Effect of MDRI on house prices

As a next step, we explore how household sentiment revealed by the mortgage default risk index (MDRI) impacts house prices. Specifically, we use different lags of the MDRI to examine its impact at different time horizon. And to account for mortgage market conditions, we add as regressors two variables that we constructed from HMDA data: the growth rate of total amount of mortgage lending in the previous year (*Loansum*), and the percentage of mortgage loans that are classified as subprime (*Subprime*).

According to the results reported in Table 3.3, an increase in the MDRI index lowers house price growth in the following three to six months. In the regression in which all lags are included (see model 8), the coefficients for the lags between three and six months are statistically significant and range between -0.00017 and -0.0012. Further, as anticipated we

Table 3.3: Predictive power of MDRI for the house price appreciation rates (ΔHP).

	House Price appreciation rate (ΔHP)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ΔMDRI_{t-1}	-0.0003						-0.0003	-0.0003
	(-0.81)						(-0.92)	(-0.84)
ΔMDRI_{t-2}		0.0005					0.0002	-0.0001
		(1.47)					(0.46)	(-0.16)
ΔMDRI_{t-3}			-0.0008***				-0.0008**	-0.0013***
			(-2.58)				(-2.41)	(-3.79)
ΔMDRI_{t-4}				-0.0009***				-0.0017***
				(-2.75)				(-4.74)
ΔMDRI_{t-5}					-0.0003			-0.0013***
					(-1.01)			(-3.62)
ΔMDRI_{t-6}						-0.0008**		-0.0012***
						(-2.53)		(-3.70)
ΔHP_{t-1}	0.8355***	0.8355***	0.8353***	0.8352***	0.8354***	0.8352***	0.8353***	0.8338***
	(213.11)	(213.15)	(213.11)	(213.05)	(213.03)	(213.01)	(213.03)	(212.32)
HF_{t-1}	-0.0002***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***
	(-7.74)	(-7.77)	(-7.77)	(-7.80)	(-7.76)	(-7.79)	(-7.77)	(-7.97)
Subprime_{t-12}	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***
	(-12.68)	(-12.72)	(-12.69)	(-12.63)	(-12.68)	(-12.66)	(-12.71)	(-12.63)
$\Delta\text{Loan supply}_{t-12}$	0.0005***	0.0005***	0.0005***	0.0005***	0.0005***	0.0005***	0.0005***	0.0005***
	(5.19)	(5.13)	(5.18)	(5.12)	(5.16)	(5.14)	(5.16)	(4.83)
Recourse	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(-3.36)	(-3.41)	(-3.32)	(-3.33)	(-3.35)	(-3.32)	(-3.30)	(-2.98)
Constant	-0.0008***	-0.0008***	-0.0008***	-0.0008***	-0.0008***	-0.0008***	-0.0008***	-0.0009***
	(-4.96)	(-4.91)	(-5.02)	(-4.96)	(-4.96)	(-4.99)	(-5.05)	(-5.42)
Number of Obs	19,254	19,254	19,254	19,254	19,254	19,254	19,254	19,254
Adjusted R-squared	0.7696	0.7697	0.7697	0.7697	0.7696	0.7697	0.7697	0.7701

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

find that the amount of mortgage credit that flows into the area serves to increase home values, while subprime lending in the previous year dampens home values in the current year.

In Appendix B, we present regression results for the 2007-2012 and the 2013-2016 subsamples (see Table B2, Panel A and Panel B, respectively), and we find that the predictive power of MDRI applies mostly for the first subsample that includes the subprime mortgage crisis. In addition to considering actual growth rates of the house price tiers, we also analyse the decomposition of house prices into their fundamental and bubble components. We find that the MDRI dampens both the fundamental component (see Table B3 in Appendix B) and the bubble component (see Table B4 in Appendix B) of house prices.

The regression results also indicate that foreclosure legislation has a significant effect on home value appreciation rates. In particular, house price growth is on average lower in the metropolitan areas located in recourse states where lenders can pursue a deficiency judgment against borrowers. The coefficient for the recourse dummy variable in Table 3.3 is statistically significant at conventional levels and equals -0.0002 across all specifications. One possible explanation is that buying a home with a mortgage is less attractive to borrowers in a recourse state where contracts are lacking the put option associated with mortgage default.

3.5.3 Effect of MDRI on foreclosure rates

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

In Table 3.4 we present results for search behaviour with lags between 1 and 6 months. We find that an increase in the MDRI lowers foreclosures for horizons between two and six months.[28] These coefficients are statistically significant and range between -0.0425 and -0.1645 (see model 8).

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

Table 3.4: Predictive power of MDRI for the Homes Foreclosed (monthly lags).

	Homes Foreclosed (HF)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ΔMDRI_{t-1}	0.0369**						-0.0200	-0.0185
	(2.04)						(-1.04)	(-0.96)
ΔMDRI_{t-2}		-0.1272***					-0.1611***	-0.1645***
		(-6.97)					(-8.05)	(-8.04)
ΔMDRI_{t-3}			-0.0587***				-0.1046***	-0.1167***
			(-3.21)				(-5.40)	(-5.63)
ΔMDRI_{t-4}				0.0181				-0.0549***
				(0.99)				(-2.66)
ΔMDRI_{t-5}					-0.0006			-0.0425**
					(-0.04)			(-2.10)
ΔMDRI_{t-6}						-0.0931***		-0.0969***
						(-5.14)		(-5.04)
HF_{t-1}	0.9729***	0.9730***	0.9729***	0.9730***	0.9729***	0.9728***	0.9730***	0.9727***
	(502.71)	(503.35)	(502.77)	(502.63)	(502.63)	(502.90)	(503.68)	(503.57)
ΔHP_{t-1}	-1.1015***	-1.1235***	-1.1190***	-1.1026***	-1.1082***	-1.1407***	-1.1508***	-1.2188***
	(-4.80)	(-4.90)	(-4.88)	(-4.80)	(-4.83)	(-4.97)	(-5.02)	(-5.31)
Subprime_{t-12}	0.0093***	0.0097***	0.0093***	0.0093***	0.0093***	0.0093***	0.0097***	0.0097***
	(6.68)	(6.93)	(6.64)	(6.65)	(6.66)	(6.69)	(6.95)	(7.01)
$\Delta\text{Loan supply}_{t-12}$	-0.0508***	-0.0492***	-0.0509***	-0.0506***	-0.0508***	-0.0515***	-0.0489***	-0.0505***
	(-8.30)	(-8.04)	(-8.31)	(-8.27)	(-8.29)	(-8.41)	(-8.00)	(-8.25)
Recourse	-0.0102***	-0.0093**	-0.0097***	-0.0100***	-0.0100***	-0.0095**	-0.0085**	-0.0076**
	(-2.70)	(-2.47)	(-2.58)	(-2.66)	(-2.65)	(-2.53)	(-2.27)	(-2.02)
Constant	0.0787***	0.0771***	0.0767***	0.0780***	0.0778***	0.0767***	0.0744***	0.0718***
	(8.42)	(8.28)	(8.21)	(8.35)	(8.33)	(8.22)	(7.97)	(7.67)
Number of Obs	19,148	19,148	19,148	19,148	19,148	19,148	19,148	19,148
Adjusted R-squared	0.9425	0.9426	0.9425	0.9425	0.9425	0.9425	0.9427	0.9428

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

And we also aggregate the Google searches for periods of three months and considers regressions with lags of up to a year. The results are presented in Table 3.5. We find also for this setting that Google searches reduce foreclosures (the coefficients for the lags between 1 and 3 months and lags between 9 and 12 months are statistically significant). As a robustness check, in Table B5 presented in Appendix B we consider the 2007-2012 and the 2013-2016 subsamples and largely find an inverse relationship between MDRI index and future foreclosures.

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

3.5.4 Effect of house prices on foreclosure rates

In Table 3.6 we report results for several alternative specifications. We disaggregate house prices to their fundamental and bubble components and study the contributing effect of these two components on foreclosures.

The finding that is robust across all specifications is that the foreclosures respond to changes in fundamental values whereby a one percent drop in fundamental home values increases the log of the homes foreclosed by 1.2, i.e. about 3.3 extra foreclosures in the following month for every 10,000 homes. The bubble component, on the other hand, does not appear to have an impact on the proportion of defaulting homeowners. We interpret this as further evidence for strategic sophistication by homeowners. A shock to the fundamental component of home values would have a long-term effect on future house prices while a shock to the bubble component would disappear over time as home values revert to their long-run equilibrium. Indeed, note that the speed of adjustment coefficient in the fundamental equation reported in Table 3.2 is significant and has the correct sign. We further examine strategic default behaviour by constructing a dummy variable for a house price declines of more than 5 percent in the past twelve months (PriceDecline>5%) and find that the foreclosures increase more in the MSAs sustaining such declines.

Table 3.5: Predictive power of MDRI for the Homes Foreclosed (three-month lags).

	Homes Foreclosed (HF)					
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta\text{MDRI}_{t-4\ to\ t-1}$	-0.0888***				-0.0966***	-0.1011***
	(-6.29)				(-6.78)	(-7.05)
$\Delta\text{MDRI}_{t-7\ to\ t-4}$		-0.0455***			-0.0582***	-0.0665***
		(-3.24)			(-4.12)	(-4.62)
$\Delta\text{MDRI}_{t-10\ to\ t-7}$			-0.0158			-0.0387***
			(-1.12)			(-2.68)
$\Delta\text{MDRI}_{t-13\ to\ t-10}$				-0.0212		-0.0301**
				(-1.50)		(-2.11)
HF_{t-1}	0.9730***	0.9727***	0.9728***	0.9728***	0.9727***	0.9722***
	(503.21)	(502.44)	(501.81)	(501.82)	(503.03)	(500.71)
ΔHP_{t-1}	-1.1509***	-1.1537***	-1.1279***	-1.1179***	-1.2132***	-1.2866***
	(-5.02)	(-5.02)	(-4.90)	(-4.87)	(-5.28)	(-5.57)
Subprime_{t-12}	0.0094***	0.0093***	0.0093***	0.0095***	0.0095***	0.0100***
	(6.77)	(6.69)	(6.70)	(6.79)	(6.82)	(7.12)
$\Delta\text{Loan supply}_{t-12}$	-0.0497***	-0.0520***	-0.0513***	-0.0523***	-0.0512***	-0.0548***
	(-8.13)	(-8.48)	(-8.36)	(-8.43)	(-8.35)	(-8.77)
Recourse	-0.0086**	-0.0094**	-0.0098***	-0.0098***	-0.0078**	-0.0070*
	(-2.29)	(-2.49)	(-2.60)	(-2.60)	(-2.06)	(-1.85)
Constant	0.0735***	0.0761***	0.0776***	0.0784***	0.0709***	0.0707***
	(7.87)	(8.14)	(8.32)	(8.39)	(7.58)	(7.54)
Number of Obs	19,148	19,148	19,148	19,148	19,148	19,148
Adjusted R-squared	0.9426	0.9425	0.9425	0.9425	0.9426	0.9427

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

Table 3.6: Predictive power of house price declines for the Homes Foreclosed.

	Homes Foreclosed (HF)			
	(1)	(2)	(3)	(4)
ΔMDRI_{t-1}	-0.0170	-0.0159	-0.0171	-0.0160
	(-0.88)	(-0.83)	(-0.89)	(-0.83)
ΔMDRI_{t-2}	-0.1634***	-0.1622***	-0.1635***	-0.1622***
	(-7.99)	(-7.92)	(-7.99)	(-7.92)
ΔMDRI_{t-3}	-0.1169***	-0.1157***	-0.1171***	-0.1158***
	(-5.64)	(-5.57)	(-5.64)	(-5.58)
ΔMDRI_{t-4}	-0.0559***	-0.0549***	-0.0561***	-0.0550***
	(-2.71)	(-2.66)	(-2.72)	(-2.66)
ΔMDRI_{t-5}	-0.0434**	-0.0426**	-0.0436**	-0.0427**
	(-2.15)	(-2.10)	(-2.15)	(-2.11)
ΔMDRI_{t-6}	-0.0987***	-0.0980***	-0.0989***	-0.0981***
	(-5.14)	(-5.10)	(-5.15)	(-5.10)
HF_{t-1}	0.9704***	0.9695***	0.9703***	0.9695***
	(467.36)	(457.02)	(466.87)	(456.89)
$\Delta\text{Fundamental}_{t-1}$	-1.1974***	-1.1980***	-1.2151***	-1.2081***
	(-4.16)	(-4.10)	(-4.17)	(-4.08)
Bubble_{t-1}	-0.0894**	-0.0862*	-0.0898**	-0.0863*
	(-2.02)	(-1.94)	(-2.03)	(-1.94)
Subprime_{t-12}	0.0090***	0.0092***	0.0089***	0.0092***
	(6.32)	(6.39)	(6.23)	(6.32)
$\Delta\text{Loan supply}_{t-12}$	-0.0480***	-0.0480***	-0.0479***	-0.0479***
	(-7.76)	(-7.75)	(-7.74)	(-7.74)
Price Decline>5%	0.0146***	0.0203***	0.0128*	0.0191**
	(2.59)	(2.75)	(1.81)	(2.20)
Recourse		-0.0024		-0.0024
		(-0.55)		(-0.55)
Recourse * Price Decline>5%		-0.0146		-0.0143
		(-1.55)		(-1.51)
Toptier			0.0003	0.0004
			(0.07)	(0.10)
Toptier * Price Decline>5%			0.0045	0.0026
			(0.49)	(0.28)
Constant	0.0626***	0.0667***	0.0621***	0.0662***
	(7.20)	(6.84)	(6.96)	(6.65)
Number of Obs	19,148	19,148	19,148	19,148
Adjusted R-squared	0.9428	0.9428	0.9428	0.9428

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

We do not find evidence that foreclosures are driven primarily by option-theoretic defaults as the coefficients for both the Recourse dummy and the interaction term of the Recourse dummy with the PriceDecline>5% dummy is not significant. These findings correspond to reported results in the recent empirical literature documenting the relative rarity of defaults due solely to strategic motives, and the relative importance of affordability constraints. Bhutta et al. (2017) find that homeowners do not walk away from their investments unless they are substantially more 'underwater' than option-theoretical models would predict, while Gerardi et al. (2018) find that default is primarily driven by income shocks rather than strategic motives.

Furthermore, the effect on foreclosures is about the same regardless of whether the decline has originated in the top tier or the bottom tier of the local housing market.

3.6 Conclusion

As of 2020, Google commands more than 92 percent of the search engine market share worldwide with an estimated number of approximately 2 trillion global searches per year (www.hubspot.com). Extant research has established that Google searches provide timely indicators for social and economic activity in a variety of domains ranging from automotive sales to the spread of infections, to asset returns in financial and housing markets. Da et al. (2011) find that search volume data predict stock returns and conclude that "search data has the potential to objectively and directly reveal to empiricists the underlying belief of an entire population of households". Chauvet et al. (2016) constructed a mortgage default risk index from data on Google search volumes for keywords such as "mortgage help" and "foreclosures assistance" and demonstrated that this index has predictive power for the returns on housing and mortgage-related assets.

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

That is, through online searches for “foreclosure help” and “mortgage assistance” households can access relevant information that helps them avert foreclosures.

We also report new results on the interaction between housing and mortgage markets which suggest some degree of household strategic behaviour. In particular, we find that declines in the fundamental component of house prices lead to an increase in foreclosure rates while declines in the transitory component of house prices have no statistically significant effect.

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

Chapter 4

Information transmission vs. information learning via Google search

4.1 Introduction

Most of the literature in the field of finance relies on economic data collected from actual economic activities. These data are typically observed with a time delay. Due to the delay in data collection, these data are unfortunately unsuitable for forecasting purposes. In the search for variables with predictive power for future activities and outcomes, some studies have turned to using internet search data, which are more timely and widely covered. As online searches reveal users' interests, the analysis of search data provides a possibility to predict the actual economic activity. For this purpose, appropriate query terms need to be chosen. Online search data has been shown to predict economic activity in many domains, including job search (Baker and Fradkin, 2017), investor attention (Da et al., 2011), and mortgage default risk (Chauvet et al., 2016). In this case, the internet search activity is regarded as an information disclosure process. The query terms reveal the interest of internet users and their intention to perform activities related to the search terms.

However, internet users are not only showing their interest or attention when they are searching but also collecting information in this process, which can be used by the user for decision-making. For example, the internet has become an information source for patients to look for treatments or check their doctors' advice (Orgaz-Molina et al., 2015; Ziebland et al., 2004). After the outbreak of the Covid-19 pandemic, the search intensity for the query term "COVID-19 treatment" increased dramatically and still shows a high correlation with the number of infections in the following two years (see Figure 4.1). Studies in information retrieval have quantified the knowledge obtained during search sessions (Hersh et al., 2002; Gadiraju et al., 2018). In addition, literature shows that online searches affect the decision-making process of Internet users (Roscoe et al., 2016). In this process, an online search is not only an information disclosure process but also a learning process.

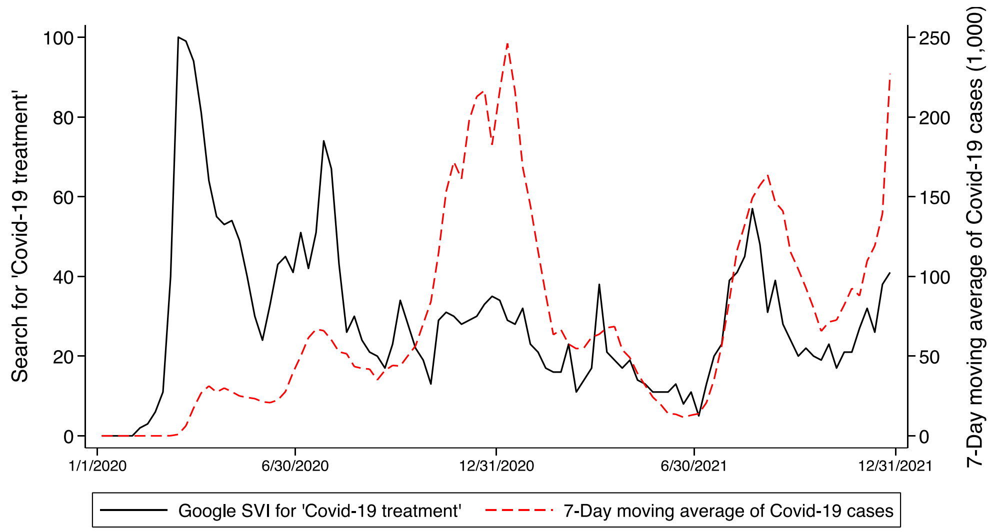


Figure 4.1: National search for "Covid-19 treatment" and the number of Covid-19 cases.

Notes: This figure depicts the dynamics of the original Google Search Volume Index (SVI) data for the term "Covid-19 treatment" and the 7-day moving average of the Covid-19 cases in the US.

The combination of information disclosure and learning during online searches raises new concerns about how online search data can be used to predict actual economic activities, e.g., mortgage default. On the one hand, if the information disclosure process works, online household searches for mortgage defaults would indicate higher default risk. On the other hand, if the information-learning process works, online searches can help internet users avoid default and hence be associated with a lower default risk. These conflicting mechanisms lead to different predictions about the effect of online search data on mortgage default risk. Therefore, in this study, we will examine the overall impact of Google search on mortgage default risk by considering the information disclosure and the learning effect of online searches.

Furthermore, the choice of query terms is highly affected by the possession of relevant knowledge in the search topic and can further affect the search efficiency in finding helpful information. Studies in information retrieval have shown that search engine users tend to use broader terms at the beginning of search sessions due to a lack of prior domain knowledge in areas related to the search topic (Vakkari et al., 2003). As they learn about the topic, they will search for more specific query terms (Wildemuth, 2004). Based on different assumptions regarding the possession of relevant knowledge in the search topic, this study defines two kinds of search activities, i.e., naïve search activity and sophisticated search activity, distinguished by the query terms used in Google searches. Specifically, naïve search activity refers to the search behaviour of borrowers lacking pertinent information, with the associated query terms indicating help-seeking actions related to mortgage default; the sophisticated search activity refers to the search patterns of borrowers who have relevant knowledge about viable solutions to retain their homes when in default and use those specific solutions as their search queries.

To separate the conflicting information disclosure and learning processes, we examine the effects of Google search on mortgage default performance in the short and long term within the recent four quarters. The empirical results show that sophisticated search activity has a positive impact on the percentage change in mortgages being in 90+ days delinquency, in line with the result of Chauvet et al. (2016) that the mortgage default risk index derived from the Google search for mortgage default help shows predictive power on mortgage delinquency indicators. However, the results also show that sophisticated search activity postpones foreclosure starts in the long term, which implies that borrowers can learn from their online searches and take action to avoid losing their homes. In comparison, it is also shown that naïve search activity positively impacts foreclosure starts in the short term. The conflicting effects of Google searches on mortgage delinquency and foreclosure starts support the hypothesis that the Google search activity is a combination of information disclosure and learning processes. However, due to the delay in the information learning process, the online search activity of households is more likely to be positively (negatively) related to the mortgage default performance of households in the short (long) term. The above findings are robust in alternative settings that take into consideration loan supply characteristics, financial literacy of households, alternative measure of mortgage delinquency rate, alternative calculation method of abnormal Google SVI, and the variation of Google SVI data at different time points.

Furthermore, it is shown that the impacts are less significant in states that did not experience a substantial house price drop in the recent four quarters. The results also suggest that sophisticated Google searches help households decrease the risk of mortgages within 90+ days of delinquency entering the foreclosure process.

This study contributes to the literature in the real estate field and the use of internet search data in several aspects. First, unlike previous studies in finance, which mainly regard online searches as an information disclosure process, this study provides supporting empirical evidence that the online searches of households are also an information-learning process. Second, the results suggest that the two processes play a relatively dominant role in the short and long term. Specifically, the information disclosure (learning) process is more likely to dominate in the relatively short (long) term. Third, the results show that the information-learning effect of online searches in real estate can be affected by the choice of query terms. Simple online searches for mortgage default help may not provide enough helpful information and further help borrowers avoid mortgage delinquency or foreclosure starts.

The remainder of this study is organized as follows. Section 4.2 reviews the literature on the current use of internet search data in finance and economics and discusses the hypothesis development. Section 4.3 introduces options provided by Fannie Mae to delinquent borrowers to keep their homes. Section 4.4 explains the construction method of the two measures of online search activities and introduces other variables used in the empirical section. The empirical results of the study are presented in Section 4.5, and Section 4.6 concludes.

4.2 Literature review

4.2.1 Use of internet searches as an economic indicator

As previously noted, the online searches conducted by internet users reflect their interests related to the search topic. This can be viewed as an inadvertent disclosure of information, offering a theoretical foundation for studies that utilize Google search data to gauge actual economic activities. McLaren and Shanbhogue (2011) state that internet search data provide a timely indicator for various economic activities. They find the data useful in predicting unemployment and house prices in the United Kingdom. Baker and Fradkin (2017) construct the Google Job Search Index based on the Google Trends data for terms containing the word "jobs". They use it to measure the overall job search activity and show that the index is correlated with the job search statistics from the comScore web panel and the American Time Use Survey. Other studies also show that Google search data can help to predict actual economic activities, such as the price volatility for energy commodities, crude oil prices, and oil demand and consumption (Afkhami et al., 2017; Li et al., 2015; Yu et al., 2019).

Internet search data has also been used to study investor sentiment and attention in the asset market. Based on the search volume data from searches that use the stock ticker or company name of stocks in the Russell 3000 index as query terms, Da et al. (2011) construct a new measure of retail investor attention. In a later study by Da et al. (2015), they use the search volume data for a set of query terms related to household concerns (e.g., recession, unemployment, and bankruptcy) to construct a market-level measure of investor sentiment for the U.S. stock market. In comparison, Gao et al. (2020) construct an investor sentiment index for 38 countries based on the search volume data for two sets of search terms that are either related or unrelated to economics and finance. Their results suggest the index works well as a contrarian predictor of country-level stock market returns.

In real estate research, studies also use Google searches to measure demand for houses for sale or rent. Beracha and Wintoki (2013) use the search intensity for terms related to real estate to measure the housing demand change for a particular city. According to their results, the search volume data predict the abnormal house price change in the city relative to the overall U.S. housing market. Similarly, Wu and Brynjolfsson (2015) demonstrate that Google search data is useful in predicting house prices. Further, instead of using the internet search data for terms related to real estate, van Dijk and Francke (2018) use the number of online listed properties and the number of clicks on those properties to create a house market tightness indicator and show that it has predictive power on both house prices and housing market liquidity. In their recent study, Aroul et al. (2022) construct a housing market negative sentiment index based on search volume data for specific real estate and economic terms from the 20 cities covered by the Case-Shiller house price index and find that the negative sentiment index reduced house price returns.

This study contributes to the literature on using online search data to measure the mortgage default risk of households. Compared with other mortgage default risk measures that are based on ex-post loan-level delinquency or foreclosure data, this new measure provides a real-time predictor for potential mortgage default risk. Webb (2009) shows that the Google search volume for the term "foreclosure" is highly correlated with the actual U.S. home foreclosures over the period from 2005 to 2009, which may provide an early warning system for home foreclosure. Askitas and Zimmermann (2011) show that the weekly search volume for "hardship letter" relates well to the 30-day delinquency rate for prime mortgage loans, and the searches for other query terms, such as "short sale", "REO" and "FHA", also relate well to housing market tensions.

To our knowledge, this research is most closely related to Chauvet et al. (2016). In their study, Chauvet et al. (2016) construct a mortgage default risk index based on the search volume for terms reflecting the assistance-seeking behaviour of households for mortgage default or foreclosure, such as "mortgage default help" and "foreclosure help". They show that the new default risk measure helps to predict housing returns, mortgage delinquencies, and the premiums of subprime credit default swaps. However, they only regard Google searches as an information disclosure process and do not consider the possible information-learning effect of online searches. In comparison, this study examines the overall effect of Google searches on mortgage default while taking into account both the information disclosure effect and the information-learning effect of online searches. Furthermore, this study makes a step towards examining the effect of different search terms on the predictive power of Google search data on mortgage default.

4.2.2 Internet search as a learning process

While the online search activity provides a relatively objective reflection of the interest of internet users, the users are not searching online aimlessly. Instead, internet users often use web searches to acquire new knowledge and satisfy learning-related objectives, which is also referred to as ‘search as learning’ in information retrieval.

Studies in information retrieval have examined the knowledge obtained from information search sessions. Hersh et al. (1995, 2002) compare the correct rate of users answering questions before and after using information retrieval systems and find an increase in the correct rate after searching in the information retrieval system. More recently, Gadiraju et al. (2018) use a formulated knowledge test to quantify the knowledge gained by users before and after internet search sessions on the web. They find an average increase of almost 20 percent in knowledge gained among about 70 percent of the users. Eickhoff et al. (2014) study the evolution of query terms within search sessions and also find evidence of knowledge gained both within a single search session and across sessions. Further, by asking the participants to answer ill-defined questions, Illies and Reiter-Palmon (2004) find that participants' information search activity helps provide more original and more appropriate answers to the questions.

Other studies have also investigated possible factors that can affect search efficiency, including individual expertise in using the internet, expertise in solving information problems, domain knowledge, problem complexity, among others (Arguello et al., 2012; Brand-Gruwel et al., 2005; Lei et al., 2013; Walhout et al., 2017; Wirth et al., 2016). Many studies have also emphasized the importance of prior knowledge of the internet user in specific areas related to the search topics, i.e., domain knowledge, for search efficiency (Sanchiz et al., 2017a; Sanchiz et al., 2017b). Monchaux et al. (2015) compare the search performance between psychology students and students from other disciplines when searching for psychology information from a given website and find that the former group outperforms the latter. Sanchiz et al. (2017a) state that prior domain knowledge improves the search efficiency of older adults with respect to website navigation and the production and reformulation of query terms.

Nevertheless, to our knowledge, there has yet to be a paper studying the search as a learning phenomenon in the housing market. The study of Damianov et al. (2021) provides some evidence in line with the search as learning phenomenon. The searches of households for query terms related to mortgage default help or foreclosure help reduce their default risk at the market level, implying that households may learn from their online searches and use the information to avoid foreclosure. This study makes a further step to examine the possible influencing factors of the information-learning effect of online searches of households.

Specifically, the search activities of households regarding mortgage default are divided into naïve and sophisticated groups based on different search terms used in the information search sessions. This study examines and compares the usefulness of the two kinds of searches in helping households avoid mortgage delinquency or keep their houses after being in delinquency.

4.2.3 Hypothesis development

According to previous literature, the online search activity of households related to mortgage default is likely to affect their default performance through the information disclosure and learning processes.

On one hand, the online searches of households for query terms related to mortgage default shows their concern regarding mortgage delinquency (Chauvet et al., 2016), which is an information disclosure process and suggests a higher mortgage default and foreclosure risk of households. In this case, a positive relationship between online search activity and mortgage default risk of households is expected. Conversely, through the information disclosure process, due to the learning outcomes from online searches, households might discover ways to prevent further delinquency and foreclosure, leading to a negative correlation. The two conflicting effects make the overall effect of online searches less predictable, which can be either positive or negative.

However, while online searches instantly capture the immediate default concerns of households, there is a delay in finding and acting up actionable information and ultimately resolving the mortgage default issue. For example, excluding the preparation time for a mortgage modification application, it typically takes 30 to 90 days to finish the approval process. Therefore, the overall effect of online searches on mortgage default is more likely to be dominated by the information disclosure (learning) process in the short (long) term. With the assumption that households can act on the information from their online searches to avoid future mortgage delinquency or foreclosure, online searches are expected to show a positive (negative) impact on mortgage default in the short (long) term.

Hypothesis 1*: The online search activity of households regarding mortgage default is a combination of information disclosure and learning processes.*

Hypothesis 2*: The online search activity of households is more likely to be positively (negatively) related to the mortgage default performance of households in the short (long) term.*

Prior research has highlighted the importance of pre-existing domain knowledge in enhancing search efficiency, which is documented that participants with prior domain knowledge in the search-related area perform better than those without the knowledge during information searching (Monchaux et al., 2015; Sanchiz et al., 2017a). This may be attributed to the help of prior domain knowledge in selecting appropriate query terms. According to a study by Vakkari et al. (2003) regarding the information search activity of students for the preparation of a research proposal, students tend to use broader search terms at the beginning of their search due to the lack of domain knowledge about the research topic. Similarly, Wildemuth (2004) also finds that to solve given clinical problems with the help of a factual database, medical students tend to narrow the query terms during the search process by adding search concepts iteratively. Nordlie (1999) states that a common feature of the search queries used at the beginning of the search session is too general in relation to the intention of the user. In short, the choice of query terms can reflect the possession of relevant information related to the search task of the internet user, illustrated by the reformulation of query terms during the search process.

For the search activity of households regarding mortgage default, they are also likely to start from general terms and reformulate their query terms to incorporate information related to mortgage default solutions newly obtained during their search processes. The reformulation will increase the search efficiency in finding useful information. Intuitively, the search for query terms directly linked to feasible mortgage default solutions is most likely to provide the information that can help households get out of mortgage delinquency or foreclosure. Hence, we propose the following hypothesis:

Hypothesis 3*: Online searches using query terms more (less) related to mortgage default solutions are more likely to have a negative (positive) association with mortgage delinquency and foreclosure.*

Another cause of a possible negative correlation between mortgage default and online searches is the pre-existing financial knowledge of households regarding mortgage default solutions. It might be the case that the online search activity is conducted by financially literate households verifying their existing knowledge regarding mortgage default solutions instead of learning from the internet. The negative relationship between online searches and mortgage default is actually due to the negative relationship between the financial literacy of households and their mortgage default risk. Lusardi and Mitchell (2014) emphasize the importance of financial literacy in economic decision-making, such as making retirement plans or investment decisions (Hastings and Tejeda-Ashton, 2008; Lusardi and Mitchell, 2007). It is also found that borrowers from the financial industry are less likely to default (Agarwal et al., 2017). Therefore, it is more likely to find a significant negative relationship between online search and mortgage default in areas with a higher financial literacy level. Consequently, we propose the following hypothesis:

Hypothesis 4: *The potential negative relationship between online search and mortgage default is more significant in states with a higher financial literacy level.*

4.3 Options to avert foreclosure

During the 2007 subprime crisis, one of the main concerns of the U.S. government was how to help delinquent borrowers keep their homes. If the delinquent borrowers cannot catch up on their mortgage payments, a common outcome for the borrowers is the loss of their homes through either short sale, deed-in-lieu, or foreclosure. However, the loss of their houses not only hurts borrowers but also slows down the recovery of house prices and the economy (Campbell et al., 2011). Foreclosed houses also bring negative externalities to the neighbourhood due to poor maintenance and lead to other societal problems, such as an increase in crime within nearby areas (Arnio et al., 2012; Cui and Walsh, 2015) and a decline in the physical and mental health of households (Houle, 2014; Libman et al., 2012). These outcomes encourage borrowers to take action to avoid default (or catch up on mortgage payments to avert foreclosure). The following are options provided by Fannie Mae for borrowers who are struggling to make their mortgage payments but still want to keep their homes: [29]

Mortgage refinance: A mortgage refinance replaces the existing mortgage with a new loan, ideally with a lower interest rate. The new mortgage can also differ in length and/or type of mortgage. For example, the new interest rate can be lower than the original one, which makes the monthly mortgage payment more affordable. However, the application for a mortgage refinance has relatively high requirements for the borrower, for example, no missed mortgage payments, sufficient home equity, and a relatively low debt-to-income ratio. Borrowers may also apply for a mortgage refinance due to the decrease in mortgage interest rates in the market, even if they are not forced to do so by financial difficulties with making mortgage payments.

Forbearance: A forbearance is given by the lender that allows the borrower to pause or reduce their mortgage payments for a limited period to deal with their short-term financial difficulties. Typically, the forbearance period is 3 to 6 months, with renewal up to 12 months.

[29] Information about the options to help borrowers keep their homes can be found on the website of Fannie Mae. https://www.knowyouroptions.com/options-to-stay-in-your-home/overview.

Therefore, this is more suitable for borrowers with short-term financial hardship but is not a permanent solution to mortgage default. The borrower must repay the amount paused or reduced after the forbearance has ended. Loans in forbearance agreements are still categorized as being in delinquency.

Repayment plan: A repayment plan is an option to catch up on mortgage payments by allowing the borrower to add the past-due amount to the current mortgage payment over a specified period (e.g., 3, 6, or 9 months). This is usually used when the borrower is not eligible for refinancing or does not wish to refinance their mortgages.

Payment deferral: A deferral can solve mortgage delinquency by allowing the borrower to move the overdue mortgage payments to the end of the mortgage term. Unlike the repayment plan, the borrower will keep the current mortgage payment amount. Therefore, it is suitable for borrowers not qualifying for a repayment plan and can be used at the end of a forbearance plan.

Mortgage modification: A mortgage modification is a change to the existing mortgage terms by the lender in various respects, such as interest rate, payment amount, and length of the mortgage. A mortgage modification seeks to make monthly payments more manageable by adjusting one or multiple mortgage terms. This can include extending the loan duration, lowering interest rates, or incorporating unpaid interest into the principal balance. There are similarities between mortgage refinance and mortgage modification. However, the former has a relatively high requirement for the borrower (e.g., no missed mortgage payment), while the latter is more suitable for borrowers behind on their payments. Once the lender approves a mortgage modification agreement, the loan transitions from the default category to the current one. It is worth noting that the delinquent borrower can still apply for a mortgage modification even after they receive a foreclosure notice from the lender. They can avoid being foreclosed if the lender approves the applications.

This study uses Google search volume data for selected queries, including some of the options mentioned above, to measure the search behaviour of households. The data are downloaded from Google Trends. However, compared with Google Search, where internet users search online, Google Trends has stricter restrictions on the query term length for downloading the Google search volume data of corresponding query terms. Using all the abovementioned options is too long to formulate a joint query term. Considering the availability of Google SVI data for these options, this study only uses the term “forbearance”, “mortgage modification”, and “mortgage refinance” as part of the final joint search term. The detailed construction method of the search terms used in this study will be introduced in the next section.

4.4 Data

4.4.1 Measure of search activity

To answer the research questions, this study utilizes quarterly data from every U.S. state, spanning from the fourth quarter of 2006 to the fourth quarter of 2018. Specifically, this research employs the monthly Google SVI data between January 2006 and December 2018 from the U.S. to construct metrics representing household search behaviours related to mortgage default.[30] Our sample timeframe covers the 2007-2008 financial crisis and the subsequent recovery phase, allowing a comprehensive analysis of the relationship between the Google search behaviour of households and their mortgage default risk over the economic cycle.

For a comprehensive examination of how household online search behaviour affects mortgage default, this study categorizes and contrasts two kinds of search activities: naïve and sophisticated. These are differentiated based on whether the households have relevant information about how to avoid mortgage default and foreclosure. Specifically, the naïve search activity of the households is defined as Google searches conducted by households lacking basic information about the feasible methods, as listed in Section 4.3, to deal with mortgage delinquency and foreclosure. Due to the lack of relevant basic information, households are likelier to begin their searches using general terms related to mortgage default assistance. To be more specific, this study uses search terms that combine words including “mortgage”, “foreclosure”, “help”, and “assistance” in different ways to represent the naïve search activity of households. The detailed search terms are given in Table 4.1.[31]

In comparison, the sophisticated search activity of households is defined as the Google search activity conducted by households who know the exact solutions, as listed in Section 4.3, available to them to keep their houses when faced with mortgage default risk. Borrowers can know these solutions through previous personal experience, i.e., prior domain knowledge, or their online searches. Specifically, this study represents the sophisticated search activity of households by using searches that employ those options as query terms. Further, as sending a hardship letter to the lender is a common practice to prove financial hardship when applying for forbearance or mortgage modification, the term “hardship letter” is also used as a search

[30] Although Google provides the Google Trends data from 2004, the data before 2006 is excluded due to the low availability of the data for single search terms in this period and the extreme fluctuation of the data, which does not match the reality.

[31] It is worth noting that after the households get information about detailed methods to avoid mortgage delinquency and foreclosure, they will also revise their search terms to incorporate this information. The revised search activity is no longer defined as naïve search activity.

Table 4.1: Joint and independent search terms.

Abbreviation	Search term	Geographic regions
Naïve search activity		
USASVI1	foreclosure help	U.S.
USASVI2	mortgage help	U.S.
USASVI3	mortgage assistance	U.S.
USASVI4	mortgage foreclosure	U.S.
USASVI5	housing assistance	U.S.
ASVIN, USASVI11	foreclosure help + mortgage help + mortgage assistance + mortgage foreclosure + housing assistance	U.S. states, U.S.
Sophisticated search activity		
USASVI6	forbearance	U.S.
USASVI7	loan modification	U.S.
USASVI8	mortgage modification	U.S.
USASVI9	mortgage refinance	U.S.
USASVI10	hardship letter	U.S.
RASVIS, USASVI12	forbearance + loan modification + mortgage modification + mortgage refinance + hardship letter	U.S. states, U.S.

Notes: The column *Abbreviation* gives the label of the abnormal search volume index (ASVI) for each of the search terms, which is calculated as the 6-month moving average of the corresponding Google search volume data minus its 12-month moving average. Specifically, ASVIN and RASVIS are calculated using state-level data, and USASVI is calculated using U.S. country-level data.

term to represent the sophisticated search activity of households. The detailed search terms used to represent the sophisticated search activity of households in this study are given in Table 4.1.[32]

4.4.2 Data restriction of Google SVI

A disadvantage of Google SVI data is that data availability is restricted to some extent due to the underlying construction method of the Google Trends data. As the construction of Google Trends data for a query term is based on the corresponding search volume data for that query term, for some query terms with low search volume, their Google SVI will appear as "0" or with missing SVI data. The Google SVI data would be less instructive with too many "0" or missing values. Data availability would be further restricted with the shrink of the geographical level of the Google SVI data, for example, from the country to the state.

This study uses two methods to deal with the data availability restriction. First, according to Google Trends' guidelines, the SVI data is not affected by the order of words in a search

[32] As Google restricts the length of the search query for Google SVI data, only options with high search frequencies are used in this study. Therefore, the term "payment deferral" and "repayment plan" are excluded from the term list used in this study.

term. Additionally, the SVI for a particular search term encompasses results for its derivative terms, with additional words before or after the original search term.[33] Hence, this study employs core words related to a specific topic to formulate independent search terms, ensuring coverage of the search volume for all pertinent terms. For example, the SVI data for the term "foreclosure help" also covers the searches for terms like "help with foreclosure" or "home foreclosure help". Second, following the method of Chauvet et al. (2016), instead of using the SVI for each of the independent search terms, this study uses the SVI for joint search terms. That is, independent search terms in each group are combined to be joint search terms with a plus sign ("+"). According to Google Trends' guidelines, the SVI for a joint search term combined with a plus sign includes the searches for each independent search term within the combined joint term.[34] Compared with the SVI for an independent search term, the data for a joint search term provides a comprehensive measure of the search activity of households for all terms in relevant topics and is less affected by the data availability restriction. The final independent and joint search terms used in this research to measure the naïve and sophisticated search activities are given in Table 4.1.

Figure 4.2 shows the dynamics of the SVI for the two kinds of search activities from 2006 to 2018. Overall, both kinds of search activities increased from the beginning of the period and reached the highest point around the second quarter of 2009. From that point, the search volume dropped quickly until 2011, after which point the SVI kept falling smoothly until it reached the pre-crisis level in 2014. Specifically, compared with the naïve search activity, the sophisticated search activity of households fluctuates to a larger extent. One possible explanation is that households focus on searches that can give them useful information. They might start with naïve searches but then do more searches using terms related to sophisticated activity after they get relevant information.

It is also interesting to compare the effect of online searches for each independent search term, especially the effect of searches for feasible options that can help delinquent borrowers keep their houses. However, due to data availability, this study can only collect country-level SVI data for independent search terms. The independent search terms reflecting naïve and sophisticated search activities are also presented in Table 4.1. In Figure C1 in Appendix C, the figure in Panel A (Panel B) shows the trend of the SVI for each of the independent search terms

[33] A sample showing the influence of the order of words in a search term is provided by Google, available at: https://support.google.com/trends/answer/4359582?hl=en.
[34] A sample of joint search terms using the plus sign is provided by Google, available at: https://support.google.com/trends/answer/4359582?hl=en.

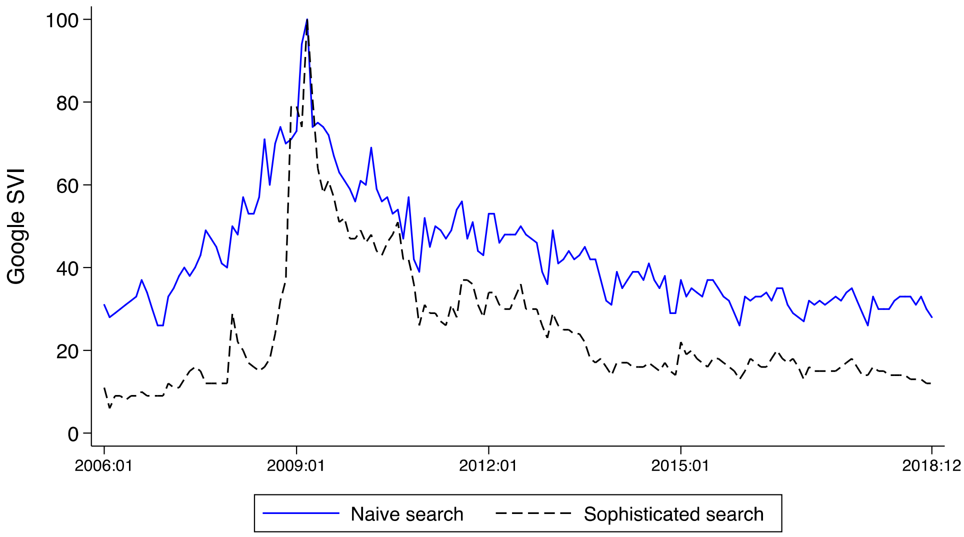


Figure 4.2: National naïve and sophisticated search activities.

Notes: This figure shows the trend of the original U.S. Google SVI data for the two joint search terms listed in Table 4.1. The blue solid line represents the dynamics of SVI for the joint search term, i.e., “foreclosure help + mortgage help + mortgage assistance + mortgage foreclosure + housing assistance”, which is used to measure naïve search activity of households. The black dashed line represents the dynamics of SVI for the joint search term, i.e., “forbearance + loan modification + mortgage modification + mortgage refinance + hardship letter”, which is used to measure the sophisticated search activity of households.

constituting the joint search term for naïve (sophisticated) search activity. Similar to the trend of the SVI for joint search terms, the SVI for most independent search terms increased quickly from the beginning and reached its highest point around 2009, then decreased until reaching the pre-crisis level. The only exception is the SVI for the term “forbearance”, which has remained high since 2009 and experienced another sudden increase followed by a decrease around 2017.

4.4.3 Calculation of Abnormal Google SVI

To deal with the extreme fluctuation of the original SVI data and decrease the impact of time trends and seasonality, this study calculates the abnormal SVI (ASVI) as the 6-month moving average of the SVI minus its 1-year moving average.[35] The calculation is as follows:

$$ASVI_{i,t} = \frac{1}{6}\sum_{j=0}^{5} SVI_{i,t-i} - \frac{1}{12}\sum_{j=0}^{11} SVI_{i,t-j} \tag{4.1}$$

where $\frac{1}{6}\sum_{i=0}^{5} SVI_{t-i}$ and $\frac{1}{12}\sum_{i=0}^{11} SVI_{t-i}$ represent the average SVI over the preceding six months and twelve months before month *t* for each geographical area *i*, which can be either a U.S. state or the entire U.S., respectively. A large positive ASVI reflects the sudden increase

[35] As a robustness check, we also use another calculation method of abnormal SVI, which calculates the abnormal SVI as the 3-month moving average of the SVI minus the 1-year moving average. Our results are robust with the new calculation method.

in household searches for information about mortgage defaults and foreclosures. The ASVI data is further standardized to be normally distributed with zero mean and unit variance. The calculation excludes the first 11 observations for each panel, which reduces the final sample period from December 2006 to December 2018. For simplicity, state-level abnormal SVI is labelled as *ASVIN* for naïve search activity and *ASVIS* for sophisticated search activity. Meanwhile, country-level abnormal SVI measures are labelled as *USASVI1, USASVI2, USASVI3, …, USASVI12* for different independent or joint search terms. The detailed labels for the abnormal SVI of different terms are presented in Table 4.1.

A special case in this study is the online search for “mortgage refinance”. Although households may use a mortgage refinance to avoid mortgage delinquency and foreclosure, they may also refinance their mortgages due to decreased mortgage interest rates in the loan market. Therefore, online searches for “mortgage refinance” might be unrelated to mortgage default. As the term is part of the joint search term reflecting the sophisticated search activity, the corresponding abnormal SVI for sophisticated search activity, i.e., ASVIS, will be less relevant to mortgage delinquency and foreclosure, which may impact the empirical results and conclusion. To control for the impact of the mortgage rate change on ASVIS, we first regress ASVIS on the change in mortgage interest rate using the following equation:

$$ASVIS_{i,t} = \sum_{j=0}^{6} \left(\theta \ \Delta Mtg30_{t-j}\right) + \delta_i + \varepsilon_{i,t} \tag{4.2}$$

where $ASVIS_{i,t}$ is the abnormal SVI reflecting the sophisticated search activity in U.S. state *i* at month *t* from Equation (4.1); $\Delta Mtg30_{t-j}$ is the change in 30-year fixed mortgage rate at month t-j; δ_i is state-fixed effect for U.S. state *i*; and $\boldsymbol{\varepsilon}_{i,t}$ denotes the vector of idiosyncratic errors. Equation (4.2) is estimated by ordinary least squares (OLS). Then we get the estimated abnormal SVI for sophisticated search activity for U.S. state *i* at month *t*, labelled as $\widehat{ASVIS}_{\iota,t}$, and calculate the residual of Equation (4.2) as follows:

$$RASVIS_{i,t} = ASVIS_{i,t} - \widehat{ASVIS}_{\iota,t} \tag{4.3}$$

where $RASVIS_{i,t}$ is the difference between the original abnormal SVI ($ASVIS_{i,t}$) for sophisticated search activity and the estimated abnormal SVI for sophisticated search activity ($\widehat{ASVIS}_{\iota,t}$) for U.S. state *i* at month *t*, respectively. $RASVIS_{i,t}$ captures the dynamics of the sophisticated search activity of households at the state level due to only mortgage default concerns of households and is not affected by the decrease of mortgage interest in the loan market. If not specified, in the following text, the sophisticated search activity of households at the U.S. state level will be measured by RASVIS, instead of ASVIS, to control for the impact of mortgage interest rate decrease on the search behaviour of households.

4.4.4 Mortgage default variables and other control variables

This study uses two mortgage default performance measures, which are the percentage of mortgages being in 90+ days of delinquency at different quarters (*DELQ*), and the percentage of mortgages entering the foreclosure process during the quarter (*FS*). The data is from the National Delinquency Survey (NDS) conducted by the Mortgage Banks Association, which is available at quarterly frequency, and is downloaded from Bloomberg. The giant database of NDS, comprising approximately 44 billion first lien loans up to the fourth quarter of 2010 with 4 million subprime loans, enables it to be a leading representative data source of mortgage default performance measures.

Precisely, the first indicator measures the percentage of mortgages falling within 90+ days of delinquency but not in the foreclosure process yet, while the second default risk indicator measures the risk of a mortgage being in default and under the foreclosure process. The main difference between the two indicators is that borrowers in the second group face a higher risk of losing their houses. Typically, lenders will start the foreclosure process when borrowers are more than 120 days late on their mortgage payments. This means that mortgages in the second group must first be in the 90+ days of delinquency bucket. Even though a loan classified under the foreclosure start category does not necessarily lead to the borrower losing their home, as they can settle their loans or obtain a mortgage modification after a foreclosure notice, initiating the foreclosure process still indicates an increased likelihood of borrowers losing their homes. As the percentage of mortgages in 90+ days of delinquency and the percentage of mortgages in foreclosure starts are a stock variable and a flow variable, respectively, this study uses the difference value of the former and the original value of the latter in regressions. We match the abnormal SVI (ASVIN and RASVIS) of the third month in each quarter with the quarterly mortgage delinquency performance data.

Figure 4.3 shows the national-level dynamics of the two mortgage default performance measures between January 2006 and December 2018. Both measures were relatively stable in the period until 2007. After that, the two measures increased significantly and reached the highest point around 2010, then turned to decrease until they reached a relatively stable level around 2016.

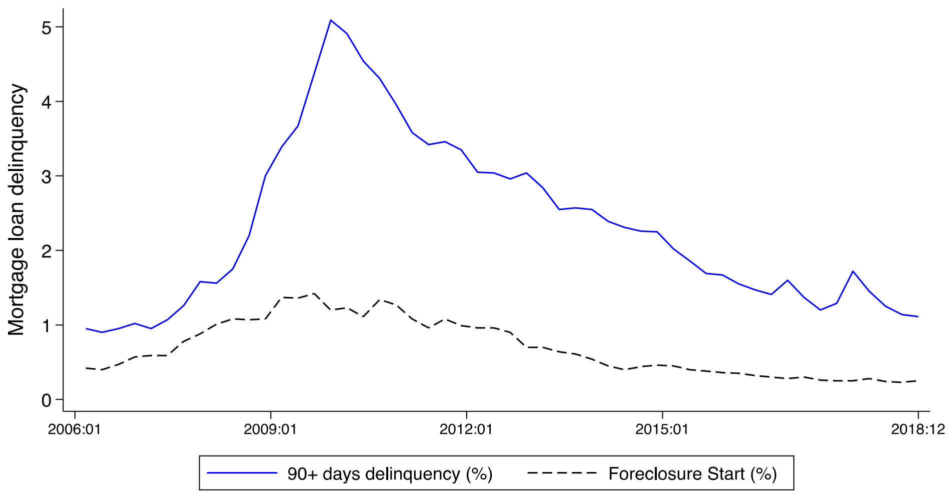


Figure 4.3: National mortgage loan delinquency.

Notes: This figure depicts the dynamics of two mortgage loan default indicators in the U.S. The blue solid line shows the movement of the percentage of mortgages in 90+ days delinquency, and the blue dashed line shows the movement of the percentage of mortgages in foreclosure starts.

To control for the impact of house prices on mortgage default risk, we use the state-level Zillow Home Value Index (*HP*) for mid-tier homes, which reflects the typical value of homes in the 35th to 65th percentile housing market ranges. Beyond house prices, this study also includes per capita personal income (*Income*) and unemployment rate (*Unemp*) as control variables to control for the macroeconomic impact on default risk. The data is sourced from the Federal Reserve Bank of St. Louis. To account for the impact of the loan market condition on the default behaviour of the borrower, similarly to the approach in Chapter 2, we include the loan supply amount (*Loansum*) and the percentage of the subprime mortgage (*Subprime*) as control variables, which are derived from the loan-level data from the Home Mortgage Disclosure Act (HMDA) from 2007 to 2018. Lastly, to control for the impact of education on the default risk of households, this study also includes the percentage of high school graduates or higher in the population by state, labelled as *Highschool_pct*, the data of which is provided by the Federal Reserve Bank of St. Louis.

Table 4.2 provides the summary statistics for the original value of all variables used in this study and the stationary test results for the variable after data transformation. The corresponding data transformation method for each variable is shown in the last column. All variables used in the final regressions are stationary after the corresponding transformation, either the first difference, logarithm, or both methods. Table 4.3 provides the correlation coefficients for the state-level measures of naïve and sophisticated search activities (i.e., ASVIN and RASVIS, respectively) and other variables. Except for the correlation coefficient between naïve search activity and foreclosure starts, which is insignificant, the other correlation

Table 4.2: Descriptive statistics.

Variables	Abbr.	N	Mean	Max	Min	Std. Dev.	Stationary test	Transformation	Geographic regions
Naïve search	USASVI1	2499	0.00	2.71	-3.50	1.00	-10.68***	original value	U.S.
	USASVI2	2499	0.00	3.82	-3.15	1.00	-13.16***	original value	U.S.
	USASVI3	2499	0.00	3.52	-3.03	1.00	-14.76***	original value	U.S.
	USASVI4	2499	0.00	2.35	-2.87	1.00	-15.85***	original value	U.S.
	USASVI5	2499	0.00	2.45	-2.79	1.00	-23.5***	original value	U.S.
	USASVI11	2499	0.00	2.84	-2.69	1.00	-11.45***	original value	U.S.
	ASVIN	2499	0.00	4.75	-4.13	1.00	-29.26***	original value	U.S. states
Sophisticated search	USASVI6	2499	0.00	3.22	-2.07	1.00	-13.38***	original value	U.S.
	USASVI7	2499	0.00	4.21	-1.65	1.00	-11.68***	original value	U.S.
	USASVI8	2499	0.00	4.27	-1.76	1.00	-10.62***	original value	U.S.
	USASVI9	2499	0.00	3.76	-3.04	1.00	-21.72***	original value	U.S.
	USASVI10	2499	0.00	3.61	-1.86	1.00	-7.48***	original value	U.S.
	USASVI12	2499	0.00	4.48	-2.17	1.00	-15.47***	original value	U.S.
	ASVIS	2499	0.00	4.81	-4.32	1.00	-27.36***	original value	U.S. states
Mortgage in 90+ days delinquency (%)	DELQ	2499	2.12	9.28	0.29	1.23	-33.34***	first-difference	U.S. states
Mortgage in foreclosure starts (%)	FS	2499	0.62	3.76	0.09	0.41	-6.95***	logarithm	U.S. states
House price ($)	HP	2499	208865	637947	87430	95285	-2.41***	log first-difference	U.S. states
Per capital personal income ($)	Income	2499	44669	83391	28422	8963	-5.13***	log first-difference	U.S. states
Unemployment rate (%)	Unemp	2499	6	15	2	2	-7.87***	first-difference	U.S. states
High school graduate or higher (%)	Highschool_pct	2499	88	94	78	3	-4.16***	original value	U.S. states
Loan supply amount (1,000$)	Loansum	2448	35458	500830	1841	55823	-8.47***	log first-difference	U.S. states
Subprime loan percentage (%)	Subprime	2448	2.59	25.27	0.03	4.48	-12.7***	first-difference	U.S. states

Notes: This table presents the summary statistics for the main variables used in the empirical sections. The second column, *Abbr.*, gives the abbreviation of each variable. Specifically, USASVI1, USASVI2, …, and USASVI12 represent the abnormal Google search for different query terms at the U.S. country level, respectively; ASVIN and RASVIS represent the abnormal Google search for different search terms at the U.S. state level, respectively. The corresponding query terms for each abbreviation are shown in Table 4.1. The stationary test is conducted using Phillips-Perron unit-root tests on the value after data transformation. The *Stationary test* column gives the value of Z-statistics from the stationary test. *, **, and *** denote the null hypothesis that all panels contain unit roots are rejected at 10%, 5%, and 1% statistical levels according to the Z-statistics from the stationary test, respectively. Column *Transformation* gives the corresponding data transformation method for each variable. It is worth noting that the first difference transformation is conducted as the quarterly difference.

Table 4.3: Correlation coefficients.

	ASVIN	RASVIS	ΔDELQ	FS	ΔHP	ΔIncome	ΔUnemp	Highschool_pct
ASVIN	1***							
RASVIS	0.437***	1***						
ΔDELQ	0.208***	0.176***	1***					
FS	0.008	-0.063***	0.161***	1***				
ΔHP	-0.169***	-0.085***	-0.357***	-0.686***	1***			
ΔIncome	-0.211***	-0.31***	-0.391***	-0.246***	0.276***	1***		
ΔUnemp	0.331***	0.394***	0.508***	0.262***	-0.467***	-0.555***	1***	
Highschool_pct	-0.055***	0.011	-0.101***	-0.382***	0.232***	0.092***	-0.128***	1***

Notes: This table presents the correlation coefficients for the main variables used in the empirical sections. This table only presents the correlation coefficients for the state-level abnormal SVI index. The correlation coefficients for country-level abnormal SVI index are presented in Table C1 in Appendix C. *, **, and *** denote the coefficient estimates are significant at 10%, 5%, and 1% statistical levels, respectively. Specifically, ASVIN and RASVIS represent the abnormal Google search for query terms reflecting naïve and sophisticated search activities, respectively; ΔDELQ and FS represent the quarterly change in the percentage of mortgages 90+ days past due and the percentage of mortgages entering into foreclosure process in the quarter, respectively; ΔHP, ΔIncome and ΔUnemp represent the quarterly change in house price return, quarterly change in per capita personal income growth rate, quarterly change in the unemployment rate, respectively; Highschool_pct represents the percentage of the population with high school degree or higher in each state.

coefficients are significantly different from zero at the 1% statistical level. Specifically, measures of search activities (ASVIN and RASVIS) are positively correlated with the change in the percentage of mortgages 90+ days past due (ΔDELQ) but negatively correlated with the percentage of mortgages entering the foreclosure process in the quarter (FS). The inconsistency of the correlation relationship between abnormal Google searches and different measures of mortgage default performance also implies the low predictability of the impact of Google searches on mortgage default risk. The correlation between other control variables, including quarterly house price growth rate (ΔHP), quarterly personal income growth rate (ΔIncome), and quarterly change in the unemployment rate (ΔUnemp), and the two default performance measures are also in line with expectations. The correlation coefficients for country-level abnormal SVI for joint and independent search terms are presented in Table C1 in Appendix C.

4.5 Empirical results

4.5.1 Baseline results

This study focuses on the effects of online searches on mortgage default within the following four quarters post the online search activity. We define the short term as up to two quarters after the search, while the long term is defined as three and four quarters after the search. For predictive analysis, we regress the mortgage default performance variables on different lags of ASVI. Considering that we are using quarterly data, we include the first and third lags of ASVI in regressions to distinguish between the short- and long-term effects of ASVI. Specifically, we use the following equation to examine the relationship between the ASVI and mortgage default performance:

$$Default_{i,t} = \sum_{j=1,3}\left(\alpha_{1,j}ASVIN_{i,t-j} + \alpha_{2,j}RASVIS_{i,t-j}\right) + \sum_{m}\beta_m Controls_{i,t-m}^{m} + Year_t + \delta_i + \varepsilon_{i,t} \quad (4.4)$$

where the subscripts i and t represent U.S. states and quarterly time points, respectively. $Default_{i,t}$ represents the dependent variable, which is one of the default performance variables, either the percentage change in mortgages in 90+ days of delinquency (i.e., $\Delta DELQ_{i,t}$), or the percentage of mortgages entering the foreclosure process in that quarter (i.e., $FS_{i,t}$). $ASVIN_{i,t-j}$ and $RASVIS_{i,t-j}$ reflect the abnormal naïve and sophisticated search activities from j quarters prior to the current quarter t, respectively. $Controls_{i,t-m}^{m}$ contains an array of control variables that include lagged dependent variable ($Dep.Var_{i,t-1}$), lagged quarterly house price returns ($\Delta HP_{i,t-1\,tot-5}$), lagged quarterly growth rate of per capita personal income ($\Delta Income_{i,t-1\ to\ t-5}$), and lagged quarterly change in the unemployment rate ($\Delta Unemp_{i,t-1\ to\ t-5}$). $Year_t$ and δ_i represent the year-fixed effect and the state-fixed effect. The coefficients of interest are α_1 and α_2, which capture the effects of the naïve and sophisticated search activities on mortgage default risk of households in the short- and long-term periods.

Table 4.4 presents the estimation results for Equation (4.4). According to the results in Column (1), only RASVIS, which measures the sophisticated search activity of households, has positive and statistically significant coefficients. Specifically, a one-unit increase in one-quarter-ahead and three-quarter-ahead RASVIS relates to a 2.8 and a 1.4 basis point increase in the change in the 90+ days mortgage delinquency rate, respectively. This indicates that the sophisticated search activity is mainly information disclosure processes for predicting the 90+ days delinquency rate.

Turning to the results in Column (4), it is observed that at the 1% statistical level, ASVIN has a positive significant coefficient at lag 1, while RASVIS has a negative significant coefficient at lag 3. According to these results, a one-unit increase of one-quarter-ahead ASVIN corresponds to a 1.5 basis point increase in the foreclosure start rate, while a one-unit increase of three-quarter-ahead RASVIS corresponds to a 1.5 basis point decrease in the foreclosure start rate.

The results support the first three hypotheses of this study. First, the positive impact of ASVIN and the negative impact of RASVIS on foreclosure starts support our hypothesis that the online search activity is a combination of information disclosure and learning processes. Second, the estimated coefficient of RASVIS on FS supports our hypothesis that the search activity is more likely to show a negative effect in the relatively long term if households can learn from their online searches. Lastly, our third hypothesis about the impact of query term choice is also supported by the significance difference in the coefficient of RASVIS on ΔDELQ and FS. As the query terms used in sophisticated search activity directly link to feasible foreclosure solutions, households are more likely to get executable information from relevant searches to avoid entering foreclosure. Therefore, sophisticated search activity is more likely to show a negative effect on foreclosure starts but may not negatively affect mortgage delinquency.

Table 4.4: Baseline results.

	ΔDELQ			FS		
	(1)	(2)	(3)	(4)	(5)	(6)
ASVIN_{t-1}	0.002	0.009		0.012***	0.012***	
	(0.432)	(1.643)		(3.252)	(3.393)	
ASVIN_{t-3}	-0.004	0.001		-0.000	-0.003	
	(-0.540)	(0.133)		(-0.017)	(-1.045)	
RASVIS_{t-1}	0.028***		0.029***	-0.005		-0.001
	(3.844)		(3.962)	(-1.480)		(-0.423)
RASVIS_{t-3}	0.014**		0.013***	-0.011***		-0.012***
	(2.606)		(2.781)	(-5.009)		(-5.221)
$\mathrm{Dep.Var}_{t-1}$	0.100***	0.097***	0.101***	0.712***	0.710***	0.708***
	(3.000)	(2.985)	(3.076)	(18.009)	(17.863)	(18.272)
$\Delta\mathrm{HP}_{t-1}$	-3.338***	-3.369***	-3.325***	-4.658***	-4.717***	-4.725***
	(-5.602)	(-5.697)	(-5.547)	(-11.214)	(-11.346)	(-11.146)
$\Delta\mathrm{Income}_{t-1}$	-3.464***	-3.652***	-3.462***	-1.280***	-1.207***	-1.304***
	(-4.242)	(-4.224)	(-4.274)	(-3.711)	(-3.496)	(-3.723)
$\Delta\mathrm{Unemp}_{t-1}$	0.029*	0.036**	0.029*	0.012	0.012	0.015
	(1.766)	(2.152)	(1.713)	(1.183)	(1.191)	(1.362)
Highschool_pct	-0.016**	-0.017**	-0.016**	-0.005	-0.005	-0.005
	(-2.330)	(-2.436)	(-2.335)	(-0.947)	(-0.901)	(-0.875)
Constant	1.599**	1.680***	1.599***	0.729	0.706	0.709
	(2.673)	(2.799)	(2.680)	(1.488)	(1.444)	(1.427)
Year & State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,346	2,346	2,346	2,346	2,346	2,346
Number of States	51	51	51	51	51	51
Adjusted R-squared	0.482	0.479	0.483	0.909	0.908	0.908

Notes: The table reports how different abnormal search activities of households affect their mortgage default performance. We run the following regression using different mortgage default performance measures, either the change in 90+ days delinquency rate (ΔDELQ) or the foreclosure start rate (FS), as the dependent variables:

$$Default_{i,t} = \sum_{j=1,3} \left(\alpha_{1,j} ASVIN_{i,t-j} + \alpha_{2,j} RASVIS_{i,t-j}\right) + \sum_{m} \beta_m Controls_{i,t-m}^{m} + Year_t + \delta_i + \varepsilon_{i,t}$$

The independent variables include two abnormal SVI indices (ASVIN and RASVIS) measuring the naïve and sophisticated search activities of households, respectively. In Columns (1) and (4), both the lags of ASVIN and RASVIS are included as independent variables. In Columns (2), (3), (5), and (6), only the lags of ASVIN or RASVIS are included as independent variables. Other independent variables include the autoregressive term of the dependent variables (Dep.Var), quarterly house price growth rates (ΔHP), quarterly personal income growth rate (ΔIncome), change in the unemployment rate (ΔUnemp), and the percentage of the population with high school degree or higher (Highschool_pct). The independent variables are included with different lags. Year-fixed effect and state-fixed effect are included in all regressions. *, **, and *** denote that the coefficient estimates are significant at 10%, 5%, and 1% statistical levels, respectively. Robust t-statistics of coefficients are presented in parentheses.

Previous studies have documented the impact of loan characteristics, such as the loan-to-value ratio and credit scoring, on mortgage default risk. For example, Amromin and Paulson (2009) find that subprime mortgages have a higher default rate than prime mortgages. Further, loan supply in the housing market is also documented as an important driver of local house prices (Favara and Imbs, 2015). The house price increase can decrease the default risk of households to some extent. Therefore, we add the annual loan supply growth rate (ΔLoansum) and the annual percentage change in subprime mortgage (ΔSubprime) in the local mortgage market as new control variables.[36] The new equation is as follows:

$$\begin{aligned} Default_{i,t} = \sum_{j=1,3} \left(\alpha_{1,j} ASVIN_{i,t-j} + \alpha_{2,j} RASVIS_{i,t-j}\right) + \alpha_3 \Delta Loansum_{i,t} \\ + \alpha_4 \Delta Subprime_{i,t} + \sum_{m} \beta_m Controls_{i,t-m}^{m} + Year_t + \delta_i + \varepsilon_{i,t} \end{aligned} \tag{4.5}$$

The estimation results for Equation (4.5) are presented in Table 4.5. According to the results, the coefficients for the growth rate of the loan supply are constantly negative and significant. This is in line with expectations, as an increase in loan supply in the housing market will lead to house price appreciation and then decrease the foreclosure risk of borrowers. In comparison, the coefficients for the percentage change in subprime mortgages are significantly negative in regressions on change in 90+ days delinquency rate (ΔDELQ), but statistically insignificant on foreclosure start rate (FS).

Furthermore, previous findings regarding the effect of online searches on mortgage default performance are not affected by mortgage loan characteristics at the market level. Specifically, in Column (1), the coefficients for RASVIS are still significantly positive at both lag 1 and lag 3, indicating an information disclosure effect of sophisticated search activity on mortgage delinquency. In comparison, in Column (4), like the results in previous regressions, ASVIN only has a positive and significant coefficient at lag 1, while RASVIS only has a negative and significant coefficient at lag 3. Overall, the results still suggest that sophisticated online searches measured by RASVIS show more information-learning effect on foreclosure starts compared with naïve online searches measured by ASVIN. Further, sophisticated online searches only show evidence of an information disclosure effect on delinquency but no evidence of an information-learning effect.

[36] The addition of a new control variable drops the observation number in regressions, as the calculation of the annual growth rate of mortgage loan supply drops the observations for 2006.

Table 4.5: Impact of mortgage loan characteristics.

	ΔDELQ			FS		
	(1)	(2)	(3)	(4)	(5)	(6)
ASVIN_{t-1}	0.000	0.008		0.013***	0.013***	
	(0.088)	(1.550)		(3.196)	(3.352)	
ASVIN_{t-3}	-0.005	-0.000		-0.000	-0.003	
	(-0.751)	(-0.021)		(-0.100)	(-1.159)	
RASVIS_{t-1}	0.031***		0.030***	-0.005		-0.001
	(4.014)		(4.128)	(-1.407)		(-0.293)
RASVIS_{t-3}	0.016***		0.014***	-0.011***		-0.012***
	(2.804)		(3.091)	(-4.761)		(-5.194)
$\mathrm{Dep.Var}_{t-1}$	0.091***	0.087***	0.092***	0.708***	0.706***	0.703***
	(2.765)	(2.765)	(2.836)	(17.199)	(17.064)	(17.450)
$\Delta\mathrm{HP}_{t-1}$	-2.418***	-2.465***	-2.404***	-4.398***	-4.462***	-4.453***
	(-3.968)	(-4.043)	(-3.934)	(-9.478)	(-9.607)	(-9.482)
$\Delta\mathrm{Income}_{t-1}$	-2.727***	-3.001***	-2.724***	-1.082***	-0.994***	-1.084***
	(-3.046)	(-3.172)	(-3.063)	(-3.210)	(-2.978)	(-3.206)
$\Delta\mathrm{Unemp}_{t-1}$	0.029	0.036**	0.028	0.013	0.013	0.016
	(1.665)	(2.023)	(1.571)	(1.307)	(1.337)	(1.471)
Highschool_pct	-0.011	-0.012	-0.011	-0.001	-0.001	-0.001
	(-1.395)	(-1.476)	(-1.422)	(-0.245)	(-0.208)	(-0.145)
$\Delta\mathrm{Loansum}_{t-12\text{ to }t}$	-0.249***	-0.247***	-0.247***	-0.085**	-0.082**	-0.091**
	(-5.069)	(-4.677)	(-5.025)	(-2.193)	(-2.167)	(-2.332)
$\Delta\mathrm{Subsum}_{t-12\text{ to }t}$	0.702***	0.758***	0.691***	0.213	0.193	0.208
	(4.026)	(4.353)	(3.919)	(1.675)	(1.509)	(1.633)
Constant	1.110*	1.198*	1.122*	0.324	0.304	0.284
	(1.693)	(1.768)	(1.716)	(0.619)	(0.580)	(0.531)
Year & State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,244	2,244	2,244	2,244	2,244	2,244
Number of States	51	51	51	51	51	51
Adjusted R-squared	0.476	0.471	0.476	0.910	0.909	0.909

Notes: The table reports how different abnormal search activities of households affect their mortgage default performance, with consideration of the impact of loan characteristics at the loan market level on the relationship. We run the following regression using different mortgage default performance measures, either the change in 90+ days delinquency rate (ΔDELQ) or the foreclosure start rate (FS), as the dependent variables:

$$Default_{i,t} = \sum_{j=1,3} \left(\alpha_{1,j} ASVIN_{i,t-j} + \alpha_{2,j} RASVIS_{i,t-j}\right) + \alpha_3 \Delta Loansum_{i,t} + \alpha_4 \Delta Subprime_{i,t} + \sum_{m} \beta_m Controls_{i,t-m}^{m} + Year_t + \delta_i + \varepsilon_{i,t}$$

The independent variables include two abnormal SVI indices (ASVIN and RASVIS) measuring the naïve and sophisticated search activities of households, respectively. In Columns (1) and (4), both the lags of ASVIN and RASVIS are included as independent variables. In Columns (2), (3), (5), and (6), only the lags of ASVIN or RASVIS are included as independent variables. Other independent variables include the autoregressive term of the dependent variables (Dep.Var), quarterly house price growth rates (ΔHP), quarterly personal income growth rate (ΔIncome), change in the unemployment rate (ΔUnemp), the percentage of the population with high school degree or higher (Highschool_pct), the 1-year growth rate of mortgage loan supply (ΔLoansum), and the 1-year percentage change in subprime mortgage loans over all mortgage loans (ΔSubprime). The independent variables are included with different lags. Year-fixed effect and state-fixed effect are included in all regressions. *, **, and *** denote that the coefficient estimates are significant at 10%, 5%, and 1% statistical levels, respectively. Robust t-statistics of coefficients are presented in parentheses.

4.5.2 Impact of substantial house price drop

The negative equity of households caused by house price declines is one of the double triggers of mortgage default. However, while the house price decline increases the default risk of households, it might also encourage them to do more online searches to find solutions to their problems. Therefore, online searches may show more supporting evidence of the information disclosure effect on mortgage default in areas that experienced a substantial house price decline. Conversely, high-frequency online searches could also provide more feasible solutions to households and help them avoid mortgage default, which may show a stronger information-learning effect. Overall, in areas with substantial drops in house prices, online searches are expected to show stronger information disclosure and learning effects on mortgage default.

Therefore, we create the substantial house price drop dummy, *SHPD*, to represent whether the house price in a state dropped by more than 5% in the preceding four quarters. Specifically, we use the following equation to examine the influence of a substantial house price drop on the impact of online searches:

$$\begin{aligned} Default_{i,t} = &\sum_{j=1,3}\left(\alpha_{1,j}ASVIN_{i,t-j} + \alpha_{2,j}RASVIS_{i,t-j}\right) \\ &+ \sum_{j=1,3}\left(\alpha_{3,j}SHPD_{i,t-j} + \alpha_{4,j}(SHPD \times ASVIN)_{i,t-j} + \alpha_{5,j}(SHPD \times RASVIS)_{i,t-j}\right) \\ &+ \sum_{m}\beta_m Controls_{i,t-m}^{m} + Year_t + \delta_i + \varepsilon_{i,t} \qquad (4.6) \end{aligned}$$

where the dummy variable for the substantial house price drop, i.e., $SHPD_{i,t-j}$, is set to be 1 when the house price in state *i* at time *t-j* drops by more than 5% in the preceding four quarters, and 0 otherwise. The two interaction terms, SHPD×ASVIN and SHPD×RASVIS, measure the difference between the impacts of naïve and sophisticated Google searches on mortgage default in states that experienced substantial house price drops compared to those in states with relatively stable house prices, respectively.

The estimated results for Equation (4.6) are presented in Table 4.6. The coefficients for ASVI, i.e., α_1 and $\alpha_2$in Equation (4.6), show the effect of online searches on mortgage default in areas where house prices have dropped by less than 5% in the previous four quarters, which are qualitatively the same as the previous results. In short, the results suggest that, in states with stable house prices, naïve search activity has an information disclosure effect on

Table 4.6: Influence of substantial house price drops.

	ΔDELQ			FS		
	(1)	(2)	(3)	(4)	(5)	(6)
ASVIN_{t-1}	0.007	0.007		0.007***	0.007***	
	(1.414)	(1.280)		(2.766)	(2.859)	
ASVIN_{t-3}	-0.011*	-0.016**		-0.002	-0.002	
	(-1.694)	(-2.651)		(-0.641)	(-0.874)	
RASVIS_{t-1}	0.013***		0.015***	-0.004		-0.003
	(2.867)		(3.208)	(-1.180)		(-0.862)
RASVIS_{t-3}	-0.010*		-0.014***	-0.006**		-0.008**
	(-1.880)		(-3.042)	(-2.413)		(-2.673)
SHPD_{t-1}	0.061**	0.066**	0.059**	0.015	0.010	0.020
	(2.664)	(2.531)	(2.533)	(0.956)	(0.678)	(1.448)
SHPD_{t-3}	-0.001	-0.015	0.002	0.039***	0.041***	0.035***
	(-0.080)	(-0.880)	(0.122)	(3.620)	(3.851)	(3.139)
$(\mathrm{SHPD}\times\mathrm{ASVIN})_{t-1}$	-0.011	0.014		0.023**	0.022***	
	(-0.847)	(1.109)		(2.276)	(2.944)	
$(\mathrm{SHPD}\times\mathrm{ASVIN})_{t-3}$	0.024*	0.077***		0.010	-0.003	
	(1.839)	(5.705)		(1.191)	(-0.382)	
$(\mathrm{SHPD}\times\mathrm{RASVIS})_{t-1}$	0.039***		0.037***	-0.003		0.011**
	(2.946)		(2.908)	(-0.419)		(2.050)
$(\mathrm{SHPD}\times\mathrm{RASVIS})_{t-3}$	0.068***		0.080***	-0.015***		-0.013***
	(6.458)		(9.787)	(-3.322)		(-2.940)
$\mathrm{Dep.Var}_{t-1}$	0.053	0.068**	0.054*	0.703***	0.700***	0.696***
	(1.627)	(2.052)	(1.730)	(15.867)	(15.723)	(16.750)
$\Delta\mathrm{HP}_{t-1}$	-2.176***	-2.115**	-2.251***	-4.101***	-4.246***	-4.281***
	(-2.765)	(-2.590)	(-2.768)	(-9.790)	(-10.340)	(-9.511)
$\Delta\mathrm{Income}_{t-1}$	-3.810***	-3.737***	-3.733***	-1.113***	-1.045***	-1.248***
	(-5.611)	(-4.804)	(-5.733)	(-3.491)	(-3.298)	(-3.650)
$\Delta\mathrm{Unemp}_{t-1}$	0.028*	0.034**	0.026*	0.011	0.010	0.015
	(1.971)	(2.297)	(1.780)	(1.117)	(1.075)	(1.403)
Highschool_pct	-0.023***	-0.022***	-0.023***	-0.006	-0.006	-0.006
	(-3.470)	(-3.257)	(-3.409)	(-1.132)	(-1.129)	(-0.987)
Constant	2.162***	2.074***	2.151***	0.822	0.820	0.776
	(3.832)	(3.626)	(3.773)	(1.674)	(1.675)	(1.540)
Year & State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,346	2,346	2,346	2,346	2,346	2,346
Number of States	51	51	51	51	51	51
Adjusted R-squared	0.505	0.494	0.504	0.910	0.910	0.909

Notes: The table reports how significant house price drops in the latest four quarters affect the impact of different abnormal search activities of households on their mortgage default performance. We run the following regression using different mortgage default performance measures, either the change in 90+ days delinquency rate (ΔDELQ) or the foreclosure start rate (FS), as the dependent variables:

$$
\begin{aligned}
Default_{i,t} = &\sum_{j=1,3}\left(\alpha_{1,j}ASVIN_{i,t-j} + \alpha_{2,j}RASVIS_{i,t-j}\right)\\
&+ \sum_{j=1,3}\left(\alpha_{3,j}SHPD_{i,t-j} + \alpha_{4,j}(SHPD \times ASVIN)_{i,t-j} + \alpha_{5,j}(SHPD \times RASVIS)_{i,t-j}\right)\\
&+ \sum_{m}\beta_m Controls^{m}_{i,t-m} + Year_t + \delta_i + \varepsilon_{i,t}
\end{aligned}
$$

The independent variables include two abnormal SVI indices (ASVIN and RASVIS) measuring the naïve and sophisticated search activities of households, respectively. A new set of independent variables is the substantial house price drop dummy (SHPD) and its interaction terms with the two abnormal SVI indices (SHPD*ASVIN and SHPD*RASVIS). The dummy variable, SHPD, equals 1 for the sample period in states where house prices dropped by more than 5% in the latest four quarters. In Columns (1) and (4), the lags of ASVIN, RASVIS, and their interaction terms with the substantial house price drop dummy are included as independent variables. In Columns (2), (3), (5), and (6), only the lags of ASVIN or RASVIS and their interaction terms with the substantial house price drop dummy are included as independent variables. Other independent variables include the autoregressive term of the dependent variables (Dep.Var), quarterly house price growth rates (ΔHP), quarterly personal income growth rate (ΔIncome), change in the unemployment rate (ΔUnemp), and the percentage of the population with high school degree or higher (Highschool_pct). The independent variables are included with different lags. Year-fixed effect and state-fixed effect are included in all regressions. *, **, and *** denote that the coefficient estimates are significant at 10%, 5%, and 1% statistical levels, respectively. Robust t-statistics of coefficients are presented in parentheses.

foreclosure. In contrast, sophisticated search activity has an information disclosure effect on mortgage delinquency and an information learning effect on foreclosure.

We are more interested in the significance and sign of the coefficients for the interaction terms, i.e., SHPD×ASVI, showing the difference between the impacts of online search activity on mortgage default performance in states with and without substantial house price drop. The coefficients of the interaction terms have the same sign as the corresponding coefficients of ASVI when the coefficients are statistically significant. Specifically, in Column (1), the coefficients for SHPD×RASVIS are significantly positive at lag 1 and lag 3, and in Column (4), the coefficient of SHPD×ASVIN (SHPD×ASVIS) is significantly positive (negative) at lag 1 (lag 3). Quantitatively, compared with that in states with stable house prices, a one-unit increase in RASVIS at one-quarter-ahead and three-quarter-ahead correspond to another 3.9 and 6.8 basis points additional increase of mortgage delinquency change in states with a substantial house price drop, but the three-quarter-ahead increase in RASVIS will decrease foreclosure by another 1.5 basis points. Regarding the impact of naïve search activity, a one-unit increase in ASVIN related to another 2.3 basis point increase in foreclosure in states with substantial house price drops. Overall, the results suggest that the naïve search activity has a stronger information disclosure effect on foreclosure, while the sophisticated search activity shows a stronger information disclosure effect on mortgage delinquency and a stronger information-learning effect on foreclosure.

4.5.3 Effect on the transfer from mortgage delinquency to foreclosure starts

In this subsection, we examine the influence of online searches on the transfer from mortgage delinquency to foreclosure starts. According to the federal regulation regarding loss mitigation procedures, unless the borrowers are more than 120 days late on their mortgage payments, lenders cannot start the foreclosure process for any judicial or non-judicial foreclosure.[37] This means that the borrower must be in 90+ days of delinquency before entering the foreclosure starts group. However, before that, borrowers can still use methods, such as mortgage forbearance and mortgage modification, to avoid the start of the foreclosure process. Therefore, online searches could help borrowers avoid entering the foreclosure process by giving them relevant information about relevant methods. To measure the impact of online searches on the transfer from mortgage delinquency to foreclosure starts, the mortgage delinquency measure, i.e., ΔDELQ, and its interaction terms with two online search activity measures, i.e., ΔDELQ×ASVIN and ΔDELQ×RASIVIS, are added as independent variables into regressions on the foreclosure start rate. The new regression equation is as follows:

$$\begin{aligned} FS_{i,t} = & \sum_{j=1,3} \left(\alpha_{1,j} ASVIN_{i,t-j} + \alpha_{2,j} RASVIS_{i,t-j}\right) \\ & + \sum_{j=1,3} \left(\alpha_{3,j} \Delta DELQ_{i,t-j} + \alpha_{4,j} (\Delta DELQ \times ASVIN)_{i,t-j} + \alpha_{5,j} (\Delta DELQ \times RASVIS)_{i,t-j}\right) \\ & + \sum_{m} \beta_m Controls_{i,t-m}^{m} + Year_t + \delta_i + \varepsilon_{i,t} \end{aligned} \tag{4.7}$$

where *FS* represents foreclosure starts, ΔDELQ is the percentage change in mortgages in 90+ days delinquent. The two interaction terms, ΔDELQ×ASVIN and ΔDELQ×RASIVIS, measure the impacts of naïve and sophisticated Google searches on the transfer of mortgage delinquency to foreclosure starts.

Table 4.7 presents the regression results of Equation (4.7). Like previous results, the coefficient for ASVIN is significant and positive at lag 1, and the coefficient for RASVIS is significant and negative at lag 3. Furthermore, according to the result, the coefficient for the interaction between ΔDELQ and RASVIS is significantly negative at lag 3. Quantitatively, a one-unit increase in three-quarter-ahead RASVIS decreases the transfer from mortgage delinquency to foreclosure starts by 6.9 base points. This means that borrowers in 90+ days of delinquency can learn from sophisticated searches and use the information to decrease the risk of entering the foreclosure process. In comparison, the naïve online search activity cannot help to avoid the transfer from mortgage delinquency to foreclosure starts and only implies a higher foreclosure risk.

[37] Relevant regulation is available on the following website: https://www.consumerfinance.gov/rules-policy/regulations/1024/41/.

Table 4.7: Effect of online searches on the transfer from mortgage delinquency to foreclosure starts.

	FS		
	(1)	(2)	(3)
ASVIN_{t-1}	0.009***	0.009***	
	(3.355)	(3.639)	
ASVIN_{t-3}	-0.002	-0.004	
	(-0.565)	(-1.516)	
RASVIS_{t-1}	-0.001		0.002
	(-0.230)		(0.699)
RASVIS_{t-3}	-0.010***		-0.011***
	(-3.280)		(-3.931)
$\Delta\mathrm{DELQ}_{t-1}$	0.149***	0.118***	0.154***
	(5.553)	(5.224)	(5.664)
$\Delta\mathrm{DELQ}_{t-3}$	0.068***	0.076***	0.065***
	(5.577)	(5.692)	(5.823)
$(\Delta\mathrm{DELQ}\times\mathrm{ASVIN})_{t-1}$	0.016	0.032*	
	(0.756)	(1.852)	
$(\Delta\mathrm{DELQ}\times\mathrm{ASVIN})_{t-3}$	0.020	-0.013	
	(1.640)	(-0.865)	
$(\Delta\mathrm{DELQ}\times\mathrm{RASVIS})_{t-1}$	-0.003		0.010
	(-0.173)		(0.831)
$(\Delta\mathrm{DELQ}\times\mathrm{RASVIS})_{t-3}$	-0.069***		-0.062***
	(-6.442)		(-6.260)
FS_{t-1}	0.759***	0.728***	0.754***
	(21.920)	(23.506)	(23.043)
$\Delta\mathrm{HP}_{t-1}$	-3.586***	-3.862***	-3.731***
	(-8.798)	(-9.907)	(-8.985)
$\Delta\mathrm{Income}_{t-1}$	-0.777**	-0.740*	-0.831**
	(-2.102)	(-1.930)	(-2.164)
$\Delta\mathrm{Unemp}_{t-1}$	-0.007	-0.005	-0.004
	(-0.919)	(-0.590)	(-0.505)
Highschool_pct	-0.003	-0.004	-0.003
	(-0.636)	(-0.820)	(-0.652)
Constant	0.517	0.622	0.524
	(1.112)	(1.342)	(1.161)
Year & State FE	Yes	Yes	Yes
Observations	2,346	2,346	2,346
Number of States	51	51	51
Adjusted R-squared	0.918	0.914	0.917

Notes: The table reports how online searches affect the transfer from 90+ days of mortgage delinquency to foreclosure starts. We run the following regression using the foreclosure start rate (FS) as the dependent variable:

$$FS_{i,t} = \sum_{j=1,3}\left(\alpha_{1,j}ASVIN_{i,t-j} + \alpha_{2,j}RASVIS_{i,t-j}\right) + \sum_{j=1,3}\left(\alpha_{3,j}\Delta DELQ_{i,t-j} + \alpha_{4,j}(\Delta DELQ \times ASVIN)_{i,t-j} + \alpha_{5,j}(\Delta DELQ \times RASVIS)_{i,t-j}\right) + \sum_{m}\beta_m Controls_{i,t-m}^{m} + Year_t + \delta_i + \varepsilon_{i,t}$$

The independent variables include two abnormal SVI indices (ASVIN and RASVIS) measuring the naïve and sophisticated search activities of households, respectively. A new set of independent variables is the change in the percentage of mortgages in 90+ days of delinquency (ΔDELQ) and its interaction terms with the two abnormal SVI indices (ΔDELQ *ASVIN and ΔDELQ *RASVIS). In Columns (1) and (4), the lags of ASVIN, RASVIS, and their interaction terms with the substantial house price drop dummy are included as independent variables. In Columns (2), (3), (5), and (6), only the lags of ASVIN or RASVIS and their interaction terms with the substantial house price drop dummy are included as independent variables. Other independent variables include the autoregressive term of the dependent variables (Dep.Var), quarterly house price growth rates (ΔHP), quarterly personal income growth rate (ΔIncome), change in the unemployment rate (ΔUnemp), and the percentage of the population with high school degree or higher (Highschool_pct). The independent variables are included with different lags. Year-fixed effect and state-fixed effect are included in all regressions. *, **, and *** denote that the coefficient estimates are significant at 10%, 5%, and 1% statistical levels, respectively. Robust t-statistics of coefficients are presented in parentheses.

4.5.4 Effect of online searches for independent search terms

In this subsection, we compare the effect of the online searches for each independent search term on mortgage default performance. As mentioned earlier, due to Google SVI data availability restrictions, the SVI data for the independent search terms is §only available at the U.S. country level. We also test the impact of country-level abnormal SVI for the two joint search terms for comparison. Specifically, USASVI1 to USASVI5 are labels for the abnormal SVI related to independent search terms reflecting naïve search activity, while USASVI6 to USASVI10 are labels for the abnormal SVI related to independent search terms reflecting sophisticated search activity. USASVI11 and USASVI12 are labels for joint search terms measuring naïve and sophisticated search activities, respectively. The detailed search terms for each abnormal SVI index are presented in Table 4.1. The following equation is used in the regressions:

$$Default_{i,t} = \sum_{j=1,3} \left(\alpha_j USASVI_{t-j}\right) + \sum_{m} \beta_m Controls_{i,t-m}^{m} + Year_t + \delta_i + \varepsilon_{i,t} \qquad (4.8)$$

where USASVI is the abnormal SVI for one of the independent or joint search terms for the entire US. We use the country-level value of USASVI for different query terms at time point *t* to measure the search activity of households for those query terms across all states.

The regression results for Equation (4.8) are presented in Table 4.8 and Table 4.9, in which the dependent variables are the mortgage delinquency (ΔDELQ) and foreclosure starts (FS), respectively. The column name in each panel shows the abnormal SVI index used in corresponding regressions. According to the results shown in Table 4.8, the impacts of country-level abnormal SVI for independent search terms on mortgage delinquency are mixed and show no constant pattern. The coefficients for different abnormal SVI indices can be either significantly positive or negative at lag 1 and lag 3. These results suggest that online search effects for different search terms on mortgage delinquency are inconstant.

Table 4.8: Effect of online searches for independent search terms on mortgage delinquency.

	Dependent Variable: ΔDELQ					
Panel A: Effect of online searches for naive search terms on mortgage delinquency						
	USASVI1	USASVI2	USASVI3	USASVI4	USASVI5	USASVI11
USASVI_{t-1}	-0.049***	-0.029***	0.033***	-0.033***	0.076***	-0.007
	(-10.018)	(-4.330)	(4.946)	(-3.827)	(7.758)	(-1.005)
USASVI_{t-3}	0.028***	0.007	0.009	-0.110***	-0.018**	-0.063***
	(3.778)	(0.854)	(1.573)	(-14.117)	(-2.250)	(-6.780)
ΔDELQ_{t-1}	0.041	0.053	0.109***	0.152***	-0.031	0.139***
	(1.245)	(1.544)	(3.121)	(4.020)	(-0.730)	(3.973)
ΔHP_{t-1}	-3.481***	-3.571***	-3.391***	-3.177***	-3.956***	-3.218***
	(-5.979)	(-6.169)	(-5.742)	(-6.399)	(-5.832)	(-5.653)
$\Delta\text{Income}_{t-1}$	-3.495***	-3.464***	-3.507***	-4.174***	-3.150***	-4.218***
	(-3.852)	(-4.011)	(-4.250)	(-4.123)	(-3.532)	(-4.546)
ΔUnemp_{t-1}	0.104***	0.074***	0.023	0.087***	0.043**	0.033*
	(5.952)	(4.142)	(1.320)	(4.800)	(2.263)	(1.691)
Highschool_pct	-0.015**	-0.016**	-0.017**	-0.014**	-0.018**	-0.016**
	(-2.060)	(-2.204)	(-2.445)	(-2.141)	(-2.335)	(-2.409)
Constant	1.532**	1.611**	1.668***	1.542**	1.682**	1.650***
	(2.447)	(2.587)	(2.767)	(2.672)	(2.526)	(2.803)
Year & State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,346	2,346	2,346	2,346	2,346	2,346
Number of States	51	51	51	51	51	51
Adjusted R-squared	0.498	0.484	0.481	0.521	0.544	0.493
Panel B: Effect of online searches for sophisticated search terms on mortgage delinquency						
	USASVI6	USASVI7	USASVI8	USASVI9	USASVI10	USASVI12
USASVI_{t-1}	-0.026***	0.053***	0.059***	-0.042***	0.130***	0.001
	(-7.484)	(6.896)	(7.720)	(-6.205)	(13.173)	(0.086)
USASVI_{t-3}	0.082***	0.042***	0.040***	-0.010*	0.031***	0.026***
	(8.344)	(7.881)	(7.318)	(-1.860)	(5.298)	(4.970)
ΔDELQ_{t-1}	0.023	0.062*	0.068**	0.046	0.078**	0.052
	(0.685)	(1.853)	(2.057)	(1.387)	(2.331)	(1.569)
ΔHP_{t-1}	-3.843***	-3.702***	-3.704***	-3.741***	-3.703***	-3.636***
	(-7.026)	(-6.190)	(-6.276)	(-6.447)	(-6.392)	(-6.183)
$\Delta\text{Income}_{t-1}$	-4.938***	-2.514***	-2.580***	-3.369***	-2.675***	-3.210***
	(-4.575)	(-3.358)	(-3.415)	(-4.005)	(-3.554)	(-3.874)
ΔUnemp_{t-1}	0.097***	0.037**	0.037**	0.064***	0.023	0.058***
	(5.500)	(2.302)	(2.344)	(3.667)	(1.466)	(3.439)
Highschool_pct	-0.016**	-0.017**	-0.017**	-0.016**	-0.017**	-0.016**
	(-2.098)	(-2.356)	(-2.359)	(-2.255)	(-2.430)	(-2.275)
Constant	1.663**	1.682***	1.674***	1.654**	1.641**	1.640**
	(2.582)	(2.696)	(2.699)	(2.641)	(2.657)	(2.625)
Year & State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,346	2,346	2,346	2,346	2,346	2,346
Number of States	51	51	51	51	51	51
Adjusted R-squared	0.510	0.490	0.490	0.487	0.508	0.483

Notes: The table relates the change in 90+ days delinquency rate (ΔDELQ) to country-level abnormal SVI indices for independent and joint search terms. We run regressions using the following equation:

$$\Delta DELQ_{i,t} = \sum_{j=1,3} \left(\alpha_j USASVI_{t-j}\right) + \sum_{m} \beta_m Controls_{i,t-m}^{m} + Year_t + \delta_i + \varepsilon_{i,t}$$

The independent variables include lags of abnormal SVI indices at the country level, USASVI, either for independent or joint search terms. The column name gives the abnormal SVI index used in each regression. Specifically, USASVI1 to USASVI5 and USASVI11 measure the naïve search activity using different search terms, while USASVI6 to USASVI10 and USASVI12 measure the sophisticated search activity using different search terms. The search terms for USASVI1 to USASVI12 are "foreclosure help", "mortgage help", "mortgage assistance", "mortgage foreclosure", "housing assistance", "forbearance", "loan modification", "mortgage modification", "mortgage refinance", "hardship letter", "foreclosure help+mortgage help+mortgage assistance+mortgage foreclosure+housing assistance", "forbearance+loan modification+mortgage modification+mortgage refinance+hardship letter", respectively. The independent variables also include a set of control variables, including the autoregressive term of the dependent variable (ΔDELQ), quarterly house price growth rates (ΔHP), quarterly personal income growth rates (ΔIncome), quarterly change in the unemployment rates (ΔUnemp), and the percentage of the population with high school degree or higher (Highschool_pct). State-fixed effect and year-fixed effect are included in all regressions. *, **, and *** denote that the coefficient estimates are significant at 10%, 5%, and 1% statistical levels, respectively. Robust t-statistics of coefficients are presented in parentheses.

Table 4.9 presents the results for the regressions on foreclosure starts, which shows a more constant pattern. In most regressions, the coefficients of USASVI are significantly positive at lag 1 and significantly negative at lag 3. This is in line with our hypothesis that the online search activity of households is more likely to be positively (negatively) related to the mortgage default performance of households in the short (long) term. Furthermore, according to the results in the last column in Panels A and B, the coefficient of USASVI is significant and positive at lag 1 in Column USASVI11 and is significant but negative at both lag 1 and lag 3 in Column USASVI12. This aligns with previous findings that naïve (sophisticated) search activity has a positive (negative) effect on foreclosure starts. Overall, the results suggest that the effect of naïve search activity on foreclosure starts is mainly an information disclosure effect in the short term and a restricted information-learning effect in the relatively long term. Meanwhile, the effect of sophisticated search activity on foreclosure starts is mainly the information-learning effect, especially in the relatively long term.

Table 4.9: Effect of online searches for independent search terms on foreclosure starts.

	Dependent Variable: FS					
Panel A: Effect of online searches for naive search terms on foreclosure starts						
	USASVI1	USASVI2	USASVI3	USASVI4	USASVI5	USASVI11
$USASVI_{t-1}$	0.020***	0.013**	0.016***	0.007	0.025***	0.027***
	(5.990)	(2.638)	(3.390)	(1.410)	(5.013)	(5.394)
$USASVI_{t-3}$	-0.015**	-0.013**	-0.009***	-0.018***	0.011**	-0.007
	(-2.466)	(-2.417)	(-2.911)	(-3.699)	(2.120)	(-1.230)
FS_{t-1}	0.716***	0.712***	0.712***	0.712***	0.711***	0.721***
	(18.445)	(18.362)	(18.304)	(18.658)	(18.941)	(19.074)
ΔHP_{t-1}	-4.709***	-4.696***	-4.725***	-4.699***	-4.812***	-4.622***
	(-11.132)	(-11.125)	(-11.188)	(-11.166)	(-11.233)	(-10.842)
$\Delta Income_{t-1}$	-1.381***	-1.540***	-1.311***	-1.393***	-0.955***	-1.285***
	(-4.082)	(-4.326)	(-3.827)	(-3.788)	(-2.733)	(-3.811)
$\Delta Unemp_{t-1}$	-0.007	-0.000	0.004	0.019*	0.011	-0.007
	(-0.667)	(-0.002)	(0.365)	(1.766)	(1.005)	(-0.600)
Highschool_pct	-0.005	-0.005	-0.005	-0.004	-0.005	-0.005
	(-0.960)	(-0.925)	(-0.900)	(-0.803)	(-0.849)	(-0.944)
Constant	0.737	0.729	0.701	0.665	0.664	0.692
	(1.509)	(1.476)	(1.425)	(1.359)	(1.356)	(1.433)
Year & State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,346	2,346	2,346	2,346	2,346	2,346
Number of States	51	51	51	51	51	51
Adjusted R-squared	0.910	0.909	0.909	0.909	0.909	0.911
Panel B: Effect of online searches for sophisticated search terms on foreclosure starts						
	USASVI6	USASVI7	USASVI8	USASVI9	USASVI10	USASVI12
$USASVI_{t-1}$	-0.006**	0.011**	0.016***	-0.019***	0.031***	-0.013**
	(-2.070)	(2.243)	(3.590)	(-4.373)	(5.631)	(-2.635)
$USASVI_{t-3}$	0.005*	-0.018***	-0.021***	-0.028***	-0.019***	-0.025***
	(1.769)	(-5.702)	(-5.814)	(-6.913)	(-5.089)	(-6.260)
FS_{t-1}	0.703***	0.710***	0.715***	0.711***	0.714***	0.710***
	(18.015)	(18.266)	(18.549)	(18.635)	(18.733)	(18.362)
ΔHP_{t-1}	-4.826***	-4.682***	-4.633***	-4.718***	-4.719***	-4.663***
	(-11.351)	(-11.195)	(-11.081)	(-11.126)	(-11.312)	(-11.101)
$\Delta Income_{t-1}$	-1.340***	-1.486***	-1.457***	-1.606***	-1.467***	-1.739***
	(-3.808)	(-4.060)	(-4.016)	(-4.519)	(-4.099)	(-4.533)
$\Delta Unemp_{t-1}$	0.021*	0.003	-0.001	0.015	0.006	0.011
	(1.968)	(0.264)	(-0.079)	(1.449)	(0.556)	(1.027)
Highschool_pct	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005
	(-0.814)	(-0.908)	(-0.921)	(-0.836)	(-0.884)	(-0.864)
Constant	0.694	0.727	0.727	0.703	0.692	0.714
	(1.395)	(1.474)	(1.485)	(1.432)	(1.410)	(1.448)
Year & State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,346	2,346	2,346	2,346	2,346	2,346
Number of States	51	51	51	51	51	51
Adjusted R-squared	0.908	0.909	0.910	0.909	0.911	0.909

Notes: The table relates the foreclosure start rate (FS) to country-level abnormal SVI indices for independent and joint search terms. We run regressions using the following equation:

$$FS_{i,t} = \sum_{j=1,3} \left(\alpha_j USASVI_{t-j}\right) + \sum_{m} \beta_m Controls_{i,t-m}^{m} + Year_t + \delta_i + \varepsilon_{i,t}$$

The independent variables include lags of abnormal SVI indices at the country level, USASVI, either for independent or joint search terms. The column name gives the abnormal SVI index used in each regression. Specifically, USASVI1 to USASVI5 and USASVI11 measure the naïve search activity using different search terms, while USASVI6 to USASVI10 and USASVI12 measure the sophisticated search activity using different search terms. The search terms for USASVI1 to USASVI12 are "foreclosure help", "mortgage help", "mortgage assistance", "mortgage foreclosure", "housing assistance", "forbearance", "loan modification", "mortgage modification", "mortgage refinance", "hardship letter", "foreclosure help+mortgage help+mortgage assistance+mortgage foreclosure+housing assistance", "forbearance+loan modification+mortgage modification+mortgage refinance+hardship letter", respectively. The independent variables also include a set of control variables, including the autoregressive term of the dependent variable (Dep.Var), quarterly house price growth rates (ΔHP), quarterly personal income growth rates (ΔIncome), quarterly change in the unemployment rates (ΔUnemp), and the percentage of the population with high school degree or higher (Highschool_pct). State-fixed effect and year-fixed effect are included in all regressions. *, **, and *** denote that the coefficient estimates are significant at 10%, 5%, and 1% statistical levels, respectively. Robust t-statistics of coefficients are presented in parentheses.

4.5.5 Robustness check

4.5.5.1 Financial literacy of households

In this section, we test whether the financial literacy of households affects our findings regarding the impacts of the online search activity of households on their mortgage default performance.

Our previous results indicate a negative relationship between sophisticated search activity of households and foreclosure starts in the long term, which can imply an information learning effect of online search activity. However, the inverse correlation between online searches and mortgage default may be due to the reduced mortgage default risk associated with higher household financial literacy. Online search activity could predominantly be carried out by financially literate households seeking to confirm their pre-existing knowledge about mortgage default solutions instead of learning from online searches. Overall, the impact of online searches on mortgage default performance may be more significant for more financially literate households.

To test the possible impact of financial literacy, we use the data from the National Financial Capability Study conducted every three years since 2009 by the FINRA Foundation to construct the measure of financial literacy at the U.S. state level. The original data is available for different groups in each state categorized by age/gender, ethnicity, and education.[38] Specifically, if the values of each group indicate they can correctly answer the

[38] The data of the National Financial Capability Study is available at: https://finrafoundation.org/knowledge-we-gain-share/nfcs/data-and-downloads.

following questions, they are given one financial literacy point for each correct answer, which is then summed together to be the final points of each group.

M6: Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow?

M7: Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, how much would you be able to buy with the money in this account?

M8: If interest rates rise, what will typically happen to bond prices?

M9: A 15-year mortgage typically requires higher monthly payments than a 30-year mortgage, but the total interest paid over the life of the loan will be less.

M10: Buying a single company's stock usually provides a safer return than a stock mutual fund.

M31: Suppose you owe $1,000 on a loan and the interest rate you are charged is 20% per year compounded annually. If you didn't pay anything off, at this interest rate, how many years would it take for the amount you owe to double?

The final financial literacy points of each group are weighted using the given weights to match the Census distribution in each state. We then calculate the average value of the weighted final points in each state to measure the financial literacy level at the state level.

To separate the U.S. states into less and more literate states, we calculate and compare the financial literacy points in each state in 2009 with the average value of the points for all states that year. We create the high state financial literacy level dummy, *HFL*, to represent whether the financial literacy level in a state is higher or lower than the country's average level in 2009. Specifically, the following equation is used to examine whether the online search activity in less and more financially literate states show different impacts on mortgage default performance:

$$\begin{aligned} Default_{i,t} = {} & \sum_{j=1,3} \left(\alpha_{1,j} ASVIN_{i,t-j} + \alpha_{2,j} RASVIS_{i,t-j}\right) + \alpha_3 HFL_i \\ & + \sum_{j=1,3} \left(\alpha_{4,j} HFL_i \times ASVIN_{i,t-j} + \alpha_{5,j} HFL_i \times RASVIS_{i,t-j}\right) \\ & + \sum_m \beta_m Controls_{i,t-m}^m + Year_t + \varepsilon_{i,t} \qquad (4.9) \end{aligned}$$

where the dummy variable for the high financial literacy state group, i.e., HFL_i, is set to be 1 if the financial literacy point in state i is higher than the country's average point, and 0 otherwise.

The estimated results for Equation (4.9) are presented in Column *Full* of Table 4.10. The coefficients for ASVI, i.e., α_1 and $\alpha_2$in Equation (4.9), show the effect of online searches on mortgage default in less financially literate states, which are qualitatively the same as the previous results. The two interaction terms, HFL×ASVIN and HFL×RASVIS, measure the difference between the impacts of naïve and sophisticated Google searches on mortgage default in more financially literate states vs in less financially literate states. However, it is shown that the coefficients for the interaction terms are not statistically significant, indicating no significant influences of financial literacy on the impact of online searches on mortgage default.

We also separately run regressions on data for less and more financially literate states without the interaction terms based on Equation (4.4), the results of which are presented in Column *Less Literacy* and Column *More Literacy* of Table 4.10. According to the results, the significance level and sign of the ASVI coefficients are essentially consistent whether the data comes from states with lower or higher financial literacy.

Overall, our results suggest that the financial literacy of households has no significant influence on the impact of online searches on mortgage default performance. There is no evidence supporting the fourth hypothesis of us. The negative impact of online searches on foreclosure starts found in previous results is due to the information learning effect of online searches instead of the pre-existing knowledge of the households regarding relevant mortgage default solutions.

Table 4.10: Impact of household financial literacy.

	ΔDELQ			FS		
	Full	Less literacy	More Literacy	Full	Less literacy	More Literacy
ASVIN_{t-1}	0.006	0.006	0.000	0.010**	0.013***	0.015**
	(0.691)	(0.631)	(0.071)	(2.531)	(3.348)	(2.290)
ASVIN_{t-3}	-0.004	-0.006	-0.001	-0.003	0.001	0.002
	(-0.515)	(-0.617)	(-0.137)	(-0.804)	(0.288)	(0.448)
RASVIS_{t-1}	0.024***	0.023***	0.027**	-0.008	-0.008	-0.003
	(2.766)	(3.506)	(2.570)	(-1.547)	(-1.541)	(-0.796)
RASVIS_{t-3}	0.004	0.009	0.018**	-0.012**	-0.013***	-0.010***
	(0.452)	(1.016)	(2.470)	(-2.262)	(-4.431)	(-3.664)
HFL	-0.001			0.016**		
	(-0.096)			(2.573)		
$\mathrm{HFL}\times\mathrm{ASVIN}_{t-1}$	-0.005			0.007		
	(-0.413)			(1.093)		
$\mathrm{HFL}\times\mathrm{ASVIN}_{t-3}$	0.003			0.007		
	(0.250)			(1.102)		
$\mathrm{HFL}\times\mathrm{RASVIS}_{t-1}$	0.007			0.003		
	(0.664)			(0.423)		
$\mathrm{HFL}\times\mathrm{RASVIS}_{t-3}$	0.018			-0.002		
	(1.630)			(-0.305)		
$\mathrm{Dep.\,Var}_{t-1}$	0.102**	0.013	0.161***	0.815***	0.493***	0.759***
	(2.304)	(0.320)	(3.358)	(40.633)	(7.800)	(25.552)
$\Delta\mathrm{HP}_{t-1}$	-2.988***	-1.828**	-3.520***	-3.593***	-4.075***	-4.172***
	(-5.452)	(-2.588)	(-5.074)	(-9.122)	(-5.907)	(-9.243)
$\Delta\mathrm{Income}_{t-1}$	-3.169***	-4.289***	-2.574**	-1.389***	0.092	-1.653***
	(-3.578)	(-2.867)	(-2.617)	(-3.210)	(0.140)	(-4.061)
$\Delta\mathrm{Unemp}_{t-1}$	0.036**	0.013	0.053*	0.012	-0.008	0.034*
	(2.065)	(0.754)	(1.882)	(1.073)	(-0.851)	(1.863)
Highschool_pct	0.001	-0.010	-0.010	-0.006***	-0.004	-0.016*
	(0.568)	(-1.245)	(-0.653)	(-5.197)	(-0.429)	(-1.735)
Constant	0.123	1.052	1.047	0.687***	0.727	1.624*
	(0.859)	(1.567)	(0.795)	(7.189)	(0.907)	(2.025)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	2,346	1,196	1,150	2,346	1,196	1,150
Observations	-	26	25	-	26	25
Adjusted R-squared	0.481	0.462	0.504	0.924	0.861	0.934

Notes: The table reports how financial literacy affect the impact of different abnormal search activities of households on their mortgage default performance. We run the following regression using different mortgage default performance measures, either the change in 90+ days delinquency rate (ΔDELQ) or the foreclosure start rate (FS), as the dependent variables:

$$Default_{i,t} = \sum_{j=1,3} \left(\alpha_{1,j} ASVIN_{i,t-j} + \alpha_{2,j} RASVIS_{i,t-j}\right) + \alpha_3 HFL_i + \sum_{j=1,3} \left(\alpha_{4,j} HFL_i \times ASVIN_{i,t-j} + \alpha_{5,j} HFL_i \times RASVIS_{i,t-j}\right) + \sum_{m} \beta_m Controls_{i,t-m}^{m} + Year_t + \varepsilon_{i,t}$$

The independent variables include two abnormal SVI indices (ASVIN and RASVIS) measuring the naïve and sophisticated search activities of households, respectively. A new set of independent variables is the financial literacy dummy (HFL) and its interaction terms with the two abnormal SVI indices (HFL*ASVIN and HFL*RASVIS). The dummy variable, HFL, equals 1 for states where the financial literacy points are higher than the country's average point and 0 otherwise. Column *Less Literacy* and Column More Literacy present the regression results based on data from less and more financially literate states, respectively. Other independent variables include the autoregressive term of the dependent variables (Dep.Var), quarterly house price growth rates (ΔHP), quarterly personal income growth rate (ΔIncome), change in the unemployment rate (ΔUnemp), and the percentage of the population with high school degree or higher (Highschool_pct). The independent variables are included with different lags. Year-fixed effect and state-fixed effect are included in all regressions. *, **, and *** denote that the coefficient estimates are significant at 10%, 5%, and 1% statistical levels, respectively. Robust t-statistics of coefficients are presented in parentheses.

4.5.5.2 Alternative calculation method of abnormal SVI

Another robustness check is about the calculation method of abnormal SVI. Our previous method of calculating the abnormal SVI is minus the 6-month moving average of the SVI by its 12-month moving average. To check the robustness of our result, we also calculate the state-level abnormal SVI as the 3-month moving average of SVI minus its 12-month moving average using the following equation:

$$NASVI_{i,t} = \frac{1}{3}\sum_{j=0}^{3} SVI_{i,t-i} - \frac{1}{12}\sum_{j=0}^{11} SVI_{i,t-j} \quad (4.10)$$

where $\frac{1}{3}\sum_{i=0}^{3} SVI_{t-i}$ and $\frac{1}{12}\sum_{i=0}^{11} SVI_{t-i}$ represent the average SVI over the preceding three months and twelve months before month t, for U.S. state i. Same as the previous labelling method, abnormal SVI is labelled as NASVIN for naïve search activity and NASVIS for sophisticated search activity. Furthermore, to control for the impact of the mortgage rate change on NASVIS, we use the method described in Section 4.4.3 to regress NASVIS on lags of the change in the 30-year fixed mortgage rate and then calculate the residual of the regression. The residual is labelled as RNASVIS.

To verify the consistency of our earlier findings, we run regressions using new abnormal SVI measures based on Equation (4.4). The results are presented in Table 4.11. According to the results, the impacts of sophisticated search activity measured by RNASVIS are the same as our previous results. Specifically, it shows that sophisticated search has an information disclosure effect on 90+ days mortgage delinquency rate, but an information learning effect on foreclosure starts rate. Although the effects of naïve search measured by NASVIN differ from our previous results, they align with our hypothesis. Specifically, NASVIN has a significant

Table 4.11: Regression with alternative abnormal SVI measure.

	ΔDELQ			FS		
	(1)	(2)	(3)	(4)	(5)	(6)
NASVIN_{t-1}	0.015***	0.016***		0.002	0.003	
	(3.028)	(3.357)		(0.790)	(0.915)	
NASVIN_{t-3}	-0.002	0.004		-0.007**	-0.012***	
	(-0.455)	(0.798)		(-2.352)	(-3.437)	
$\mathrm{RNASVIS}_{t-1}$	0.013***		0.018***	-0.006*		-0.005*
	(2.740)		(3.538)	(-1.842)		(-1.887)
$\mathrm{RNASVIS}_{t-3}$	0.027***		0.025***	-0.020***		-0.022***
	(4.537)		(4.727)	(-5.993)		(-6.120)
$\mathrm{Dep.Var}_{t-1}$	0.093***	0.102***	0.093***	0.716***	0.709***	0.714***
	(2.846)	(3.155)	(2.820)	(18.845)	(18.390)	(18.737)
$\Delta\mathrm{HP}_{t-1}$	-3.388***	-3.348***	-3.390***	-4.696***	-4.796***	-4.692***
	(-5.648)	(-5.642)	(-5.707)	(-11.127)	(-11.347)	(-11.070)
$\Delta\mathrm{Income}_{t-1}$	-3.539***	-3.613***	-3.580***	-1.240***	-1.188***	-1.268***
	(-4.197)	(-4.241)	(-4.251)	(-3.598)	(-3.447)	(-3.612)
$\Delta\mathrm{Unemp}_{t-1}$	0.033**	0.031*	0.037**	0.012	0.015	0.012
	(2.015)	(1.796)	(2.247)	(1.115)	(1.345)	(1.117)
Highschool_pct	-0.017**	-0.017**	-0.016**	-0.005	-0.005	-0.005
	(-2.402)	(-2.513)	(-2.295)	(-0.923)	(-0.857)	(-0.913)
Constant	1.642***	1.696***	1.604**	0.719	0.696	0.720
	(2.753)	(2.871)	(2.654)	(1.474)	(1.407)	(1.470)
Year & State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,346	2,346	2,346	2,346	2,346	2,346
Number of States	51	51	51	51	51	51
Adjusted R-squared	0.486	0.481	0.484	0.910	0.908	0.910

Notes: The table reports how different abnormal search activities of households affect their mortgage default performance, with consideration of the impact of loan characteristics at the loan market level on the relationship. We run the following regression using different mortgage default performance measures, either the change in 90+ days delinquency rate (ΔDELQ) or the foreclosure start rate (FS), as the dependent variables:

$$Default_{i,t} = \sum_{j=1,3}\left(\alpha_{1,j}NASVIN_{i,t-j} + \alpha_{2,j}RNASVIS_{i,t-j}\right) + \alpha_3 \Delta Loansum_{i,t} + \alpha_4 \Delta Subprime_{i,t} + \sum_{m}\beta_m Controls_{i,t-m}^{m} + Year_t + \delta_i + \varepsilon_{i,t}$$

The independent variables include two abnormal SVI indices (NASVIN and RNASVIS) measuring the naïve and sophisticated search activities of households, respectively. The two new measures are based on the difference of 3-month moving average of SVI and its 12-month moving average. In Columns (1) and (4), both the lags of NASVIN and RNASVIS are included as independent variables. In Columns (2), (3), (5), and (6), only the lags of NASVIN or RNASVIS are included as independent variables. Other independent variables include the autoregressive term of the dependent variables (FS), quarterly house price growth rates (ΔHP), quarterly personal income growth rate (ΔIncome), change in the unemployment rate (ΔUnemp), and the percentage of the population with high school degree or higher (Highschool_pct). The independent variables are included with different lags. Year-fixed effect and state-fixed effect are included in all regressions. *, **, and *** denote that the coefficient estimates are significant at 10%, 5%, and 1% statistical levels, respectively. Robust t-statistics of coefficients are presented in parentheses.

positive coefficient at lag 1 in regressions on mortgage delinquency but a significant negative coefficient at lag 3 in regressions on foreclosure starts. Both the impacts of NASVIN and RNASVIS align with our hypothesis that the online search activity of households is more likely to be positively (negatively) related to the mortgage default performance of households in the short (long) term. However, it also shows that the estimated impacts of online searches can be affected by the calculation method of abnormal SVI.

4.5.5.3 Alternative measure of mortgage delinquency

Our previous findings are based on regressions on two mortgage default performance measures, i.e., 90+ days mortgage delinquency rate and foreclosure starts rate. In this section, we test the robustness of our previous findings by testing the impacts of online search activity on the percentage of mortgages in 60-day delinquency, labelled as 60DAY-DELQ. Table 4.12 presents the new regression results based on Equation (4.4).

The results show that ASVIN and RASVIS have significant positive coefficients at lag 1 and negative ones at lag 3. This contrasts with our earlier results, which indicated only a positive effect of sophisticated search activity and no notable influence of naïve search activity on 90+ days of mortgage delinquency. However, the findings align with the initial two hypotheses, indicating that both naïve and sophisticated search activities demonstrate a short-term information disclosure impact and a long-term information learning effect on 60-day mortgage delinquency.

Table 4.12: Regressions on the change in 60-days delinquency rate.

	Δ60DAY-DELQ		
	(1)	(2)	(3)
ASVIN_{t-1}	0.006**	0.012***	
	(2.391)	(4.529)	
ASVIN_{t-3}	-0.008**	-0.014***	
	(-2.242)	(-3.711)	
RASVIS_{t-1}	0.012***		0.014***
	(3.943)		(4.571)
RASVIS_{t-3}	-0.022***		-0.026***
	(-6.196)		(-8.487)
$\Delta\text{60DAY} - \text{DELQ}_{t-1}$	-0.304***	-0.298***	-0.292***
	(-17.343)	(-16.821)	(-17.075)
ΔHP_{t-1}	-0.091	-0.218	-0.065
	(-0.604)	(-1.602)	(-0.430)
$\Delta\text{Income}_{t-1}$	-2.402***	-2.307***	-2.387***
	(-7.152)	(-6.556)	(-7.383)
ΔUnemp_{t-1}	0.009	0.017*	0.009
	(1.067)	(1.884)	(1.164)
Highschool_pct	-0.009***	-0.008***	-0.008***
	(-3.040)	(-2.943)	(-3.110)
Constant	0.910***	0.884***	0.901***
	(3.720)	(3.663)	(3.822)
Year & State FE	Yes	Yes	Yes
Observations	2,346	2,346	2,346
Number of States	51	51	51
Adjusted R-squared	0.211	0.184	0.207

Notes: The table reports how online searches affect the change in the 60-day delinquency rate. We run the following regression using the foreclosure start rate (FS) as the dependent variable:

$$\Delta\text{60DAY} - \text{DELQ}_{i,t} = \sum_{j=1,3} \left(\alpha_{1,j} ASVIN_{i,t-j} + \alpha_{2,j} RASVIS_{i,t-j}\right) + \sum_{m} \beta_m Controls_{i,t-m}^{m} + Year_t + \delta_i + \varepsilon_{i,t}$$

The independent variables include two abnormal SVI indices (ASVIN and RASVIS) measuring the naïve and sophisticated search activities of households, respectively. In Columns (1) and (4), the lags of ASVIN, RASVIS, and their interaction terms with the substantial house price drop dummy are included as independent variables. In Columns (2), (3), (5), and (6), only the lags of ASVIN or RASVIS and their interaction terms with the substantial house price drop dummy are included as independent variables. Other independent variables include the autoregressive term of the dependent variables ($\Delta\text{60DAY} - \text{DELQ}$), quarterly house price growth rates (ΔHP), quarterly personal income growth rate (ΔIncome), change in the unemployment rate (ΔUnemp), and the percentage of the population with high school degree or higher (Highschool_pct). The independent variables are included with different lags. Year-fixed effect and state-fixed effect are included in all regressions. *, **, and *** denote that the coefficient estimates are significant at 10%, 5%, and 1% statistical levels, respectively. Robust t-statistics of coefficients are presented in parentheses.

4.5.5.4 Variation of Google SVI data

The Google Search Volume Index (SVI), also known as Google Trends, is constructed based on trillions of online searches per year for different query terms conducted by users of the search engine Google. While this provides a strong data basis for the construction of Google SVI and makes it a representative measure of the online search behaviour of people, it would be quite time-consuming to use all the Google search data in calculating Google SVI. Instead, according to Google, they only use a random sample of Google searches representative of all searches in the construction of Google SVI. While this reduces the processing time, a notable downside is that the Google SVI data can vary when downloaded at different times. It is concerned that the change in Google SVI values can affect the robustness of research findings.

To deal with this concern, we calculate the average of Google SVI data downloaded at 14 different times between May 2022 and September 2023, and use the average value of Google SVI to calculate the abnormal SVI for different query terms, labelled as ASVINAVG and ASVISAVG. Furthermore, same as in the previous section, we use the method described in Section 4.4.3 to regress ASVISAVG on lags of the change in the 30-year fixed mortgage rate and then calculate the residual of the regression. The residual is labelled as RASVISAVG.

The results for regressions using the new abnormal average SVI data based on Equation (4.4) are presented in Table 4.13. According to the results, our findings regarding the impacts of naïve and sophisticated search activities on mortgage delinquency performance are robust with the new measures of abnormal SVI. The only significant difference is that the impact of sophisticated search activity on foreclosure starts in the short term. In Column (4), the coefficient of RASVISAVG is significantly negative at lag 1, while the corresponding coefficient in Table 4.5 is negative but not significant. The new results suggest a more significant information learning effect of sophisticated search on foreclosure start but are still in line with our hypothesis.

Table 4.13: Regressions on abnormal average SVI.

	ΔDELQ			FS		
	(1)	(2)	(3)	(4)	(5)	(6)
ASVINAVG_{t-1}	0.002	0.017**		0.018***	0.015***	
	(0.308)	(2.537)		(5.262)	(4.800)	
ASVINAVG_{t-3}	-0.004	0.010		-0.001	-0.006**	
	(-0.525)	(1.486)		(-0.310)	(-2.194)	
RASVISAVG_{t-1}	0.041***		0.041***	-0.009***		-0.001
	(5.915)		(5.931)	(-2.685)		(-0.344)
RASVISAVG_{t-3}	0.026***		0.024***	-0.010***		-0.012***
	(4.252)		(4.905)	(-4.696)		(-5.357)
Dep. Var_{t-1}	0.092***	0.095***	0.093***	0.716***	0.713***	0.709***
	(2.808)	(2.978)	(2.927)	(18.143)	(17.969)	(18.342)
ΔHP_{t-1}	-3.222***	-3.246***	-3.197***	-4.600***	-4.653***	-4.726***
	(-5.353)	(-5.358)	(-5.237)	(-11.301)	(-11.429)	(-11.157)
$\Delta\text{Income}_{t-1}$	-3.421***	-3.529***	-3.408***	-1.251***	-1.211***	-1.320***
	(-4.356)	(-4.173)	(-4.433)	(-3.690)	(-3.588)	(-3.718)
ΔUnemp_{t-1}	0.023	0.030*	0.023	0.009	0.008	0.016
	(1.499)	(1.886)	(1.472)	(0.987)	(0.830)	(1.468)
Highschool_pct	-0.017**	-0.017**	-0.016**	-0.005	-0.005	-0.005
	(-2.426)	(-2.455)	(-2.408)	(-0.895)	(-0.892)	(-0.864)
Constant	1.640***	1.663***	1.634***	0.697	0.699	0.698
	(2.779)	(2.815)	(2.762)	(1.425)	(1.425)	(1.418)
Year & State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,346	2,346	2,346	2,346	2,346	2,346
Number of States	51	51	51	51	51	51
Adjusted R-squared	0.488	0.480	0.488	0.909	0.909	0.908

Notes: The table reports the results for regressions using abnormal SVI from averaged SVI data. We run the following regression using different mortgage default performance measures, either the change in 90+ days delinquency rate (ΔDELQ) or the foreclosure start rate (FS), as the dependent variables:

$$Default_{i,t} = \sum_{j=1,3}\left(\alpha_{1,j}ASVINAVG_{i,t-j} + \alpha_{2,j}RASVISAVG_{i,t-j}\right) + \sum_{m}\beta_m Controls_{i,t-m}^{m} + Year_t + \delta_i + \varepsilon_{i,t}$$

The independent variables include two abnormal SVI indices (ASVINAVG and RASVISAVG) calculated based on averaged SVI data downloaded from different time points. In Columns (1) and (4), both the lags of ASVINAVG and RASVISAVG are included as independent variables. In Columns (2), (3), (5), and (6), only the lags of ASVINAVG or RASVISAVG are included as independent variables. Other independent variables include the autoregressive term of the dependent variables (Dep.Var), quarterly house price growth rates (ΔHP), quarterly personal income growth rate (ΔIncome), change in the unemployment rate (ΔUnemp), and the percentage of the population with high school degree or higher (Highschool_pct). The independent variables are included with different lags. Year-fixed effect and state-fixed effect are included in all regressions. *, **, and *** denote that the coefficient estimates are significant at 10%, 5%, and 1% statistical levels, respectively. Robust t-statistics of coefficients are presented in parentheses.

4.6 Conclusion

This study examines the effect of the Google search behaviour of households on their mortgage default risk and shows that the search activity of households is a combination of the information disclosure process and the information-learning process. This study defines two kinds of online search activities of households, i.e., naïve and sophisticated search activities, and compares their impacts on mortgage default performance. Specifically, naïve search activity is defined as the Google searches conducted by households with no basic information about the feasible mortgage default solutions, while sophisticated search activity refers to the Google search of households with that information. In practice, we use the data of the Google Search Volume Index (SVI) for two groups of query terms to reflect the naïve and sophisticated search activities of households, respectively. Empirical analyses are conducted on regressions using U.S. state-level and country-level quarterly data from 2006Q4 to 2018Q4.

According to the results from regressions using the state-level SVI of joint search terms, sophisticated searches positively impact mortgage delinquency. This finding supports the information disclosure effect of online searches and is in line with the study by Chauvet et al. (2016), which shows the predictive power of Google search for mortgage help on mortgage delinquency. Meanwhile, naïve search activity shows a positive impact on foreclosure starts in the short term, while sophisticated search activity shows a negative impact on foreclosure starts in the long term. The above results have three implications: First, online search activity is not only an information disclosure process but also an information-learning process; second, the online search activity of households is more likely to be positively (negatively) related to the mortgage default performance of households in the short (long) term; last, the information learning effect can be affected by the choice of query term, as online searches using query terms more related to mortgage default solutions are more likely to have a negative association with mortgage delinquency and foreclosure. The above results are robust in alternative settings that take into consideration loan supply characteristics, financial literacy of households, alternative measure of mortgage delinquency rate, alternative calculation method of abnormal Google SVI, and the variation of Google SVI data at different time points.

Our results also suggest that the impacts of Google searches on mortgage default are stronger in states where the house price dropped by more than 5% in the recent year. Furthermore, it is also found that the sophisticated search activity of households can help to prevent mortgages within 90+ days of delinquency from entering the foreclosure process.

It is worth noting that even though we only find a positive relationship between mortgage delinquency and sophisticated online search activity from the empirical results, it does not mean online searches ultimately cannot help the delinquency of borrowers. Possibly, some borrowers can find helpful information and use it to avoid delinquency, or the searches for some independent search terms can help households avoid delinquency. Nonetheless, the predominant influence of sophisticated online searches on mortgage delinquency remains the information disclosure effect. Another possible explanation is that sophisticated online searches help borrowers to extend the delinquency period instead of entering the foreclosure process. For example, mortgages within the forbearance period are categorized as in delinquency and have no risk of entering the foreclosure process until the end of forbearance.

This study sheds new light on using online search data in real estate. While online searches show the users' interest in a specific topic, the users are also learning from their online searches. The overall relationship between online searches and the targeted topic will be a combination of an information disclosure process and an information-learning process, which might not be guaranteed to be positive or negative. However, the overall impact is more likely to be dominated by the information disclosure (learning) effect in the relatively short (long) term. Further, this relationship depends on the choice of query terms and other relevant factors.

Chapter 5

Conclusions and future work

5.1 Concluding remarks

This thesis focuses on the relationship between mortgage default and house prices while putting additional emphasis on the use of online search data of households in predicting mortgage defaults and house prices. Most previous studies on this relationship rely on mortgage default risk indicators derived from actual loan performance data. While this ensures good reliability of the data, it is highly affected by the time delay in data collection and is also highly restricted to measuring potential default risk without other economic data. To overcome this problem, Chauvet et al. (2016) introduce the Mortgage Default Risk Index, i.e., the MDRI, which is constructed based on online search data for terms like “mortgage foreclosure” and “foreclosure help”. This thesis contributes to the literature on the use of online search data to predict mortgage default and house prices.

In Chapter 2, we conduct an extensive analysis of the nature of the bidirectional relationship between mortgage defaults and house prices in an empirical macroeconomic framework. Compared with previous studies, we take the further step of doing a more detailed analysis and comparison of the interdependent relationship across different housing market segments. Specifically, we decompose the housing markets into top- and bottom-tier markets based on the house prices in each area, and into recourse and non-recourse states based on the category of Ghent and Kudlyak (2011). Furthermore, while using the homes foreclosed rate, i.e., the HF, to measure the actual mortgage default risk, we also use the Mortgage Default Risk Index derived from household online searches for mortgage default help or foreclosure help, i.e., the MDRI, to measure the potential default risk. In line with previous studies, a negative interdependent relationship is found between mortgage default and house price. Furthermore, the results suggest that the responses of mortgage default risk indicators to house price declines are stronger in non-recourse states than in recourse states. This finding is in line with the previous result of Ghent and Kudlyak (2011) that borrowers are more likely to default in non-recourse states. Moreover, the results also suggest that the MDRI has a higher impact on top-tier house prices, while the HF shows a higher impact on bottom-tier house prices. Overall, this study provides some findings in line with the strategic default behaviour of borrowers.

In Chapter 3, we examine the impact of Google search behaviour aggregated in the MDRI on local house prices and the mortgage default risk of households. Moreover, we conduct a further analysis of the relationship between the mortgage market and house prices by decomposing the house prices. Specifically, following the method of Abraham and Hendershott (1996) and Capozza et al. (2004), we decompose house prices into their long-term equilibrium components, i.e., the fundamental house prices, and the short-term deviation part from the equilibrium, including the bubble component of house prices. The results suggest that the MDRI negatively impacts house prices, and this negative impact persists on the decomposed house price components, either on the fundamental house prices or on the bubble component of house prices. This is in line with previous studies, as the increase in the MDRI implies a higher default risk. However, it is shown that the MDRI also negatively impacts foreclosure. One possible explanation is that households can learn about how to avoid foreclosure through their online searches and actively avoid foreclosure. Furthermore, it is found that only fundamental house prices have a significant negative impact on foreclosure rates. In contrast, foreclosure rates are less sensitive to the change in the short-term deviation of house prices. Meanwhile, it is also shown that, compared with the fundamental house prices, the short-term deviation of house prices is less sensitive to changes in the MDRI. To some extent, the two findings are in line with the strategic default behaviour of households.

Chapter 4 examines the impact of the Google search behaviour of households on two kinds of mortgage default outcomes: mortgage delinquency and foreclosure starts. While previous studies use online search data to predict actual economic activity, they neglect the potential information-learning effect of online searches. In this study, we define two kinds of online search activities, i.e., naïve and sophisticated search activities, based on different assumed levels of relevant information households have about feasible mortgage default and foreclosure solutions. Two kinds of search activity measures are constructed based on the state-level Google Trends data for two joint search terms representing the naïve and sophisticated search activities, respectively. Furthermore, to separate the conflicting information disclosure and information-learning effects of online searches on mortgage default, we decompose the effect and examine it separately in the relatively short term and long term. It is found that sophisticated search activity has a positive impact on the percentage of mortgages in 90+ days of delinquency but a negative impact on foreclosure starts, while the naïve search activity only has a positive impact on foreclosure starts. This supports that the Google search activity of households is a combination of information disclosure and information-learning processes. The search activity of households not only discloses information about their mortgage default risk but also helps them to find useful information to avoid entering foreclosure and keep their homes. However, the results also suggest that the two processes play a relatively dominant role in the relatively short term and long term. Specifically, the information disclosure (learning) process is more likely to dominate in the relatively short (long) term. Moreover, households are more likely to learn from sophisticated searches. It is also found that the impacts of online searches on mortgage default are more significant in the states that experienced a substantial house price drop (more than 5%) within the recent four quarters. Furthermore, it is also shown that sophisticated search activity can help to prevent mortgages in 90+ days of delinquency entering the foreclosure starts.

5.2 Future research

The study in this thesis can be further improved in several ways. From the perspective of the construction of online search measures, the choice of query terms chosen to represent the online search behaviour of households related to mortgage defaults is quite subjective and restricted. The construction of the monthly MDRI proposed by Chauvet et al. (2016) is based on state-level Google Trends data for a joint search term consisting of 11 independent search terms. The construction of the measure for the naïve and sophisticated search activities in this thesis is based on the Google Trends data for joint search terms combining five independent query terms. In both cases, the independent query terms are chosen subjectively. Although the calculation of Google SVI for a specific query term also includes the searches for other terms derived from the specific one, it is still possible that the current search terms only capture part of the search activity of households related to mortgage default. Therefore, it is still meaningful to expand the term basis and include more relevant query terms to construct measures of the search activity of households for mortgage defaults. However, due to the restriction in terms of the query term length when downloading Google Trends data, the current method that combines independent search terms as a joint search term should be improved.

Furthermore, the entire thesis is conducted within an empirical macroeconomic framework based on market-level data and lacks a theoretical framework. Although the findings from the empirical sections are in line with theoretical expectations, such as the strategic default behaviour of borrowers, it is not enough to predict the actions of specific households facing mortgage default risk. It would be useful to combine loan-level mortgage data with the online search data of households, and make a more detailed analysis. For example, if the borrowers are in 30, 60, or 90+ days of delinquency, how will their online searches help them in applications for mortgage modification or other mortgage default help programmes? Will online searches increase the possibility of getting the mortgage modification application approved?

Moreover, in Chapter 4, to deal with the coexisting information disclosure and information-learning effects of online searches on mortgage default, we examine the effect separately in the relatively short term and long term and find that the information disclosure (information-learning) effect is more likely to take the dominant role in the relatively short term (long term). However, this method can only examine the comprehensive effects of online searches in different time horizons. It would be meaningful if further studies could be conducted to examine the independent contribution of the information disclosure effect or the information-learning effect of online searches on mortgage default in different time horizons.

Finally, one of the questions for this thesis is the data basis for the measure of the online search activity of households, considering that there are other search engines on the market in addition to Google, such as Yahoo, Baidu, etc. This thesis focuses on the U.S. housing market, where Google is a leading company in the field of online search engines and has taken the dominant role in recent years. It takes about 40% of the search engine market share in 2004 (Sullivan, 2004, cited in Visser and Weideman, 2011, p.4) and has about 80% of the search engine market share since early 2009[39] after continuous growth. Therefore, the Google Search Volume Index (SVI) has become the most commonly used online search data in empirical studies, and is also used in this thesis to measure the online search behaviour of households. The results in this thesis are to some extent specific to the U.S. as they are based on Google search data from the U.S., but they may be different in other countries. For example, the dominant search engine in China is Baidu, which provides search results from websites using Chinese. The search habits in China may be different from those in the U.S., which may affect the results. Therefore, future studies should cover other regions and more factors influencing the information disclosure and learning processes of online searches.

[39] The search engine market share data for Google is available at: https://gs.statcounter.com/search-engine-market-share#monthly-200904-202109.

Appendix A

Table A1: Variable definition.

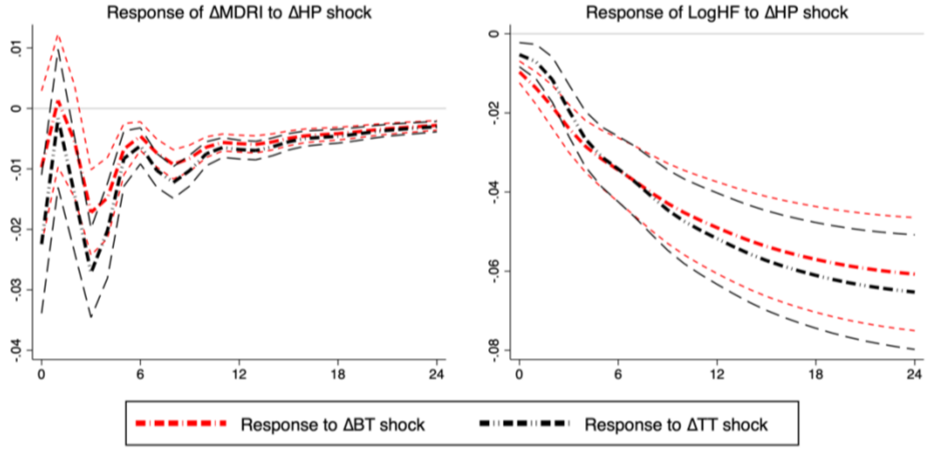
Variable	Definition
TT	**Top-tier index**: Median estimated home value for top-tier homes within a given region. Top-tier homes are homes that fall into the top tercile of home values within a given region. The data is obtained from Zillow's website.
BT	**Bottom-tier index**: Median estimated home value for all bottom-tier homes within a given region. Bottom-tier homes are the homes that fall into the bottom tercile of home values within a given region. The data is obtained from Zillow's website.
MDRI	**Mortgage Default Risk Index**: A real-time index of mortgage default risk proposed by Chauvet, Gabriel, and Lutz (2016). The index is calculated from Google search query data for terms such as "foreclosure help" and "mortgage help". It reflects households' concerns about mortgage default and possible foreclosure of their home.
HF	**Homes Foreclosed**: The number of homes (per 10,000 homes) that were foreclosed in each month. A foreclosure occurs when a homeowner loses their home to their lending institution, or when the home is sold to a third party at an auction.
Emp	**Total Nonfarm Employees**: The number of U.S. workers in the economy that excludes proprietors, private household employees, unpaid volunteers, farm employees, and the unincorporated self-employed.
Perm	**New private housing units authorized by building permits**: The permit is for a new housing unit that will be privately owned. The permits are issued by a permit-issuing authority, usually a city or town but sometimes a county covering unincorporated territory.
Indpro	**Industrial Production Index**: A measure of real output for all facilities located in the United States including manufacturing, mining, electric, and gas utilities (excluding those in U.S. territories).
Ppiaco	**Producer Price Index for All Commodities**: A measures of the average change over time in the prices received by domestic producers for their output.
Umcsent	**University of Michigan consumer sentiment:** An index provided by the University of Michigan consumer sentiment survey that measures the level of consumer confidence regarding the overall economy, based on monthly telephone surveys of around 500 consumers.
Fedfund	**Effective federal funds rate:** The interest rate at which depository institutions trade federal funds (balances held at Federal Reserve Banks) with each other overnight, which is the central interest rate in the U.S. financial market.
SP500	**S&P 500 Index:** Standard & Poor's 500 Index.

Table A2: Descriptive statistics for sub-groups.

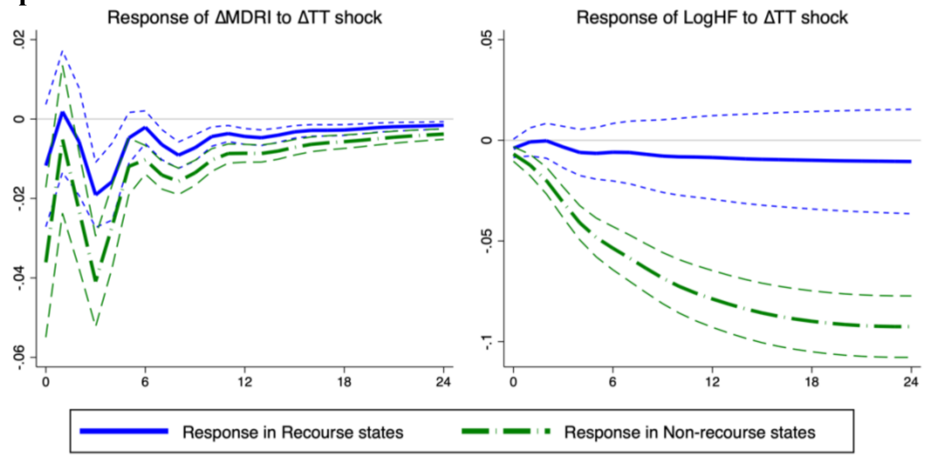
Variables	Abbr.	N	Mean	Max	Min	Std. Dev.	ADF test	Transformation	Geographic regions
Panel A: Recourse states									
Top-tier House Price (1,000$)	TT	11980	344.86	1769.90	104.50	244.75	-46.03***	Log first-difference	Metro
Bottom-tier House Price (1,000$)	BT	11792	119.90	518.70	36.90	72.58	-43.54***	Log first-difference	Metro
Mortgage Default Risk Index	MDRI	12320	129.06	554.60	24.28	64.95	-71.3***	Log first-difference	State
Homes foreclosed (%)	HF	9943	5.32	196.08	0.02	5.92	-18.57***	Logarithm	Metro
Employment (1,000)	EMP	12320	3805.87	12246.70	455.90	2425.79	-43.78***	Log first-difference	State
Building Permit	PERM	12320	2413.72	29849.58	28.14	2689.09	-71.3***	Log first-difference	State
Industrial Production Index	INDPRO	12320	97.01	103.60	84.73	4.34	-49.41***	Log first-difference	National
Producer Price Index	PPIACO	12320	183.03	208.30	141.40	18.31	-46.3***	Log first-difference	National
Consumer Sentiment	UMCSENT	12320	81.22	103.80	55.30	11.56	-71.3***	Log first-difference	National
S&P 500 Index	SP500	12320	1472.76	2384.20	735.09	392.01	-68.52***	Log first-difference	National
Federal Funds Rate	FEDFUND	12320	1.35	5.26	0.07	1.81	-11.04***	original value	National
Panel B: Non-recourse states									
Top-tier House Price (1,000$)	TT	8885	427.70	1807.80	106.00	253.60	-32.23***	Log first-difference	Metro
Bottom-tier House Price (1,000$)	BT	8735	182.76	657.30	48.40	108.26	-28.38***	Log first-difference	Metro
Mortgage Default Risk Index	MDRI	8960	113.53	627.27	13.74	65.59	-60.63***	Log first-difference	State
Homes foreclosed (%)	HF	8032	8.00	106.20	0.06	9.96	-11.6***	Logarithm	Metro
Employment (1,000)	EMP	8960	8483.27	16697.50	1448.70	6296.46	-27.86***	Log first-difference	State
Building Permit	PERM	8960	5185.28	20692.31	365.65	4405.71	-60.81***	Log first-difference	State
Industrial Production Index	INDPRO	8960	97.01	103.60	84.73	4.34	-42.13***	Log first-difference	National
Producer Price Index	PPIACO	8960	183.03	208.30	141.40	18.31	-39.49***	Log first-difference	National
Consumer Sentiment	UMCSENT	8960	81.22	103.80	55.30	11.56	-60.81***	Log first-difference	National
S&P 500 Index	SP500	8960	1472.76	2384.20	735.09	392.02	-58.44***	Log first-difference	National
Federal Funds Rate	FEDFUND	8960	1.35	5.26	0.07	1.81	-9.41***	original value	National

Notes: The table reports the descriptive statistics of all variables in the Panel VAR system. Column *ADF test* gives the value of the Z-statistics from the ADF test for the data after transformation. Specifically, the ADF test is conducted with drift and lag 1 setting. *** denote the null hypothesis that all panels contain unit roots is rejected at 1% statistical levels according to the Z-statistics and p-value from the ADF test. The last column indicates the geographical level at which the variables are observed.

Panel A. Responses to ΔTT and ΔBT shocks for the full sample



Panel B. Responses to ΔTT shock in recourse and non-recourse states



Panel C. Responses to ΔBT shock in recourse and non-recourse states

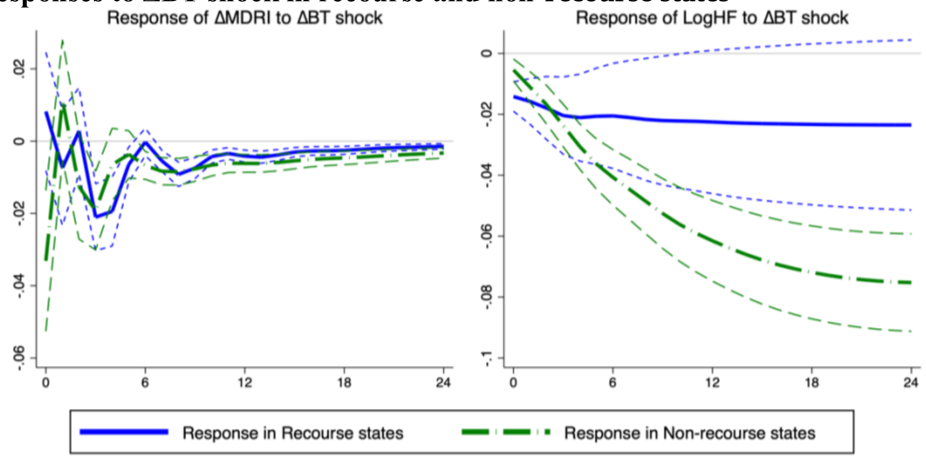
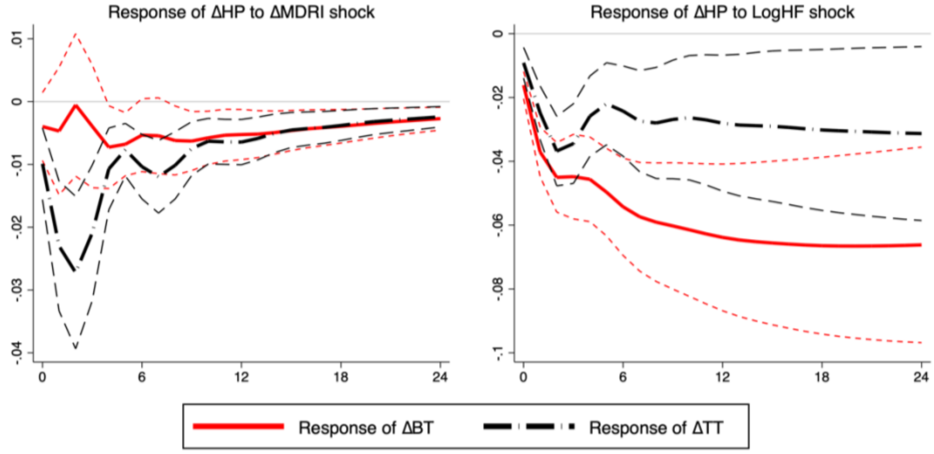


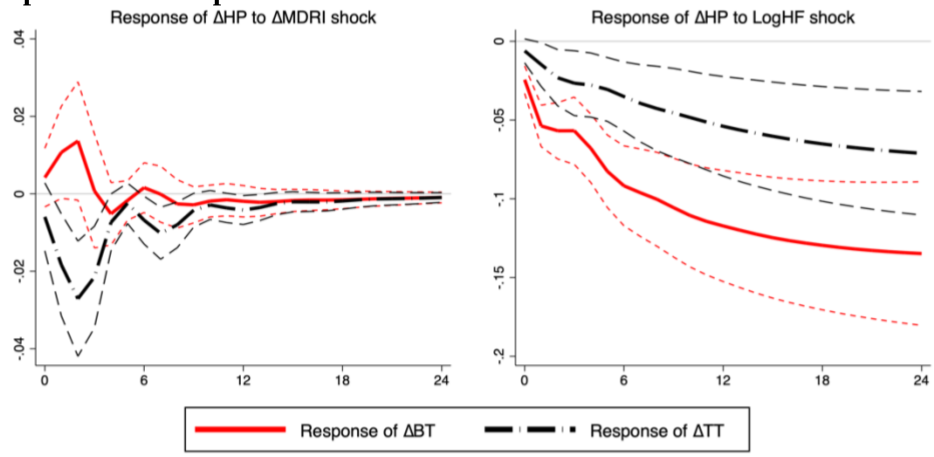
Figure A1: Standardized impulse responses of mortgage default risk to shocks to house price returns.

Notes: The thick lines represent the dynamics of the standardized impulse responses of mortgage default risk to shocks to house price returns (ΔHP) in the next 24 months. The thin lines represent the 95% confidence interval around the responses. Panel A shows the responses of mortgage default risk to shocks to house price returns in the full sample. Panel B shows the responses of mortgage default risk to a shock to top-tier house price return (ΔTT), and Panel C shows the responses of mortgage default risk to a shock to the bottom-tier house price return (ΔBT). The left and right parts of each panel show the responses of mortgage default risk that is measured by the MDRI and HF, respectively.

Panel A. Response of house price returns for the full sample



Panel B. Response of house price returns in Recourse States



Panel C. Response of house price returns in Non-recourse States

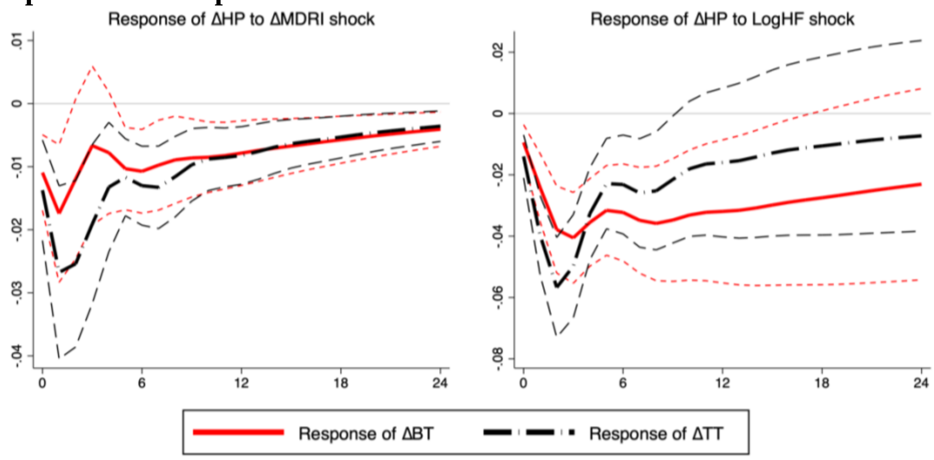


Figure A2: Standardized impulse responses of house price returns to shocks to mortgage default risk.

Notes: The thick lines represent the standardized impulse responses of house price returns (ΔHP) to shocks to the mortgage default risk in the next 24 months. The thin lines show the 95% confidence interval around the responses. Panels A, B and C show the results for the full sample, the sample of recourse states and the sample of non-recourse states, respectively. The left part and right part of each panel show house price responses when mortgage default risk is measured by the MDRI and HF, respectively.

Appendix B

Table B1: MSAs and classification of states according to state foreclosure law.

Metropolitan area	State	Recourse / Non-recourse
Akron	Ohio	Recourse
Albany	New York	Recourse
Allentown	Pennsylvania	Recourse
Atlantic City	New Jersey	Recourse
Bakersfield	California	Non-Recourse
Baltimore	Maryland	Recourse
Bellingham	Washington	Non-Recourse
Bend	Oregon	Non-Recourse
Binghamton	New York	Recourse
Bloomington	Illinois	Recourse
Boulder	Colorado	Recourse
California-Lexington Park	Maryland	Recourse
Canton	Ohio	Recourse
Charlotte	North Carolina	Non-Recourse
Chico	California	Non-Recourse
Cincinnati	Ohio	Recourse
Cleveland	Ohio	Recourse
Colorado Springs	Colorado	Recourse
Columbia	South Carolina	Recourse
Columbus	Ohio	Recourse
Corvallis	Oregon	Non-Recourse
Crestview-Fort Walton Beach-Destin	Florida	Recourse
Cumberland	Maryland	Recourse
Dallas-Fort Worth	Texas	Recourse
Dayton	Ohio	Recourse
Denver	Colorado	Recourse
Eugene	Oregon	Non-Recourse
Fayetteville	North Carolina	Non-Recourse
Flagstaff	Arizona	Non-Recourse
Fort Collins	Colorado	Recourse
Fresno	California	Non-Recourse
Glens Falls	New York	Recourse
Grand Junction	Colorado	Recourse
Green Bay	Wisconsin	Non-Recourse

(Continued on next page)

(continued)

Metropolitan area	State	Recourse / Non-recourse
Greenville	South Carolina	Recourse
Hanford	California	Non-Recourse
Hartford	Connecticut	Recourse
Johnson City	Tennessee	Recourse
Knoxville	Tennessee	Recourse
Lafayette-West Lafayette	Indiana	Recourse
Lancaster	Pennsylvania	Recourse
Las Vegas	Nevada	Recourse
Lincoln	Nebraska	Recourse
Little Rock	Arkansas	Recourse
Los Angeles-Long Beach-Anaheim	California	Non-Recourse
Madera	California	Non-Recourse
Madison	Wisconsin	Non-Recourse
Medford	Oregon	Non-Recourse
Memphis	Tennessee	Recourse
Merced	California	Non-Recourse
Milwaukee	Wisconsin	Non-Recourse
Minneapolis-St Paul	Minnesota	Non-Recourse
Mobile	Alabama	Recourse
Modesto	California	Non-Recourse
Morristown	Tennessee	Recourse
Napa	California	Non-Recourse
Nashville	Tennessee	Recourse
New Haven	Connecticut	Recourse
New London	Connecticut	Recourse
North Port-Sarasota-Bradenton	Florida	Recourse
Oklahoma City	Oklahoma	Recourse
Olympia	Washington	Non-Recourse
Philadelphia	Pennsylvania	Recourse
Phoenix	Arizona	Non-Recourse
Pittsburgh	Pennsylvania	Recourse
Pittsfield	Massachusetts	Recourse
Portland	Oregon	Non-Recourse
Prescott	Arizona	Non-Recourse
Providence	Rhode Island	Recourse
Pueblo	Colorado	Recourse
Raleigh	North Carolina	Non-Recourse
Redding	California	Non-Recourse

(Continued on next page)

(continued)

Metropolitan area	State	Recourse / Non-recourse
Reno	Nevada	Recourse
Richmond	Virginia	Recourse
Riverside	California	Non-Recourse
Sacramento	California	Non-Recourse
Salem	Oregon	Non-Recourse
Salinas	California	Non-Recourse
Salisbury	Maryland	Recourse
San Diego	California	Non-Recourse
San Francisco	California	Non-Recourse
San Jose	California	Non-Recourse
San Luis Obispo	California	Non-Recourse
Santa Cruz	California	Non-Recourse
Santa Maria-Santa Barbara	California	Non-Recourse
Santa Rosa	California	Non-Recourse
Seattle	Washington	Non-Recourse
Spartanburg	South Carolina	Recourse
Spokane	Washington	Non-Recourse
Springfield	Massachusetts	Recourse
Springfield	Ohio	Recourse
Stamford	Connecticut	Recourse
State College	Pennsylvania	Recourse
Stockton	California	Non-Recourse
Toledo	Ohio	Recourse
Tucson	Arizona	Non-Recourse
Urban Honolulu	Hawaii	Recourse
Utica	New York	Recourse
Vallejo	California	Non-Recourse
Ventura	California	Non-Recourse
Virginia Beach	Virginia	Recourse
Visalia	California	Non-Recourse
Worcester	Massachusetts	Recourse
Yakima	Washington	Non-Recourse
York	Pennsylvania	Recourse
Yuba City	California	Non-Recourse
Yuma	Arizona	Non-Recourse

Table B2: Predictive power of MDRI for the house price appreciation rates (ΔHP) at different sample period.

Panel A: 2007-2012

	ΔHP							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta MDRI_{t-1}$	-0.0003						-0.0003	-0.0001
	(-0.61)						(-0.47)	(-0.19)
$\Delta MDRI_{t-2}$		0.0011**					0.0006	0.0003
		(2.18)					(1.00)	(0.47)
$\Delta MDRI_{t-3}$			-0.0013***				-0.0011**	-0.0020***
			(-2.75)				(-2.18)	(-3.47)
$\Delta MDRI_{t-4}$				-0.0011**				-0.0025***
				(-2.41)				(-4.49)
$\Delta MDRI_{t-5}$					-0.0000			-0.0018***
					(-0.10)			(-3.30)
$\Delta MDRI_{t-6}$						-0.0014***		-0.0021***
						(-3.07)		(-4.14)
ΔHP_{t-1}	0.8216***	0.8220***	0.8214***	0.8212***	0.8217***	0.8212***	0.8216***	0.8189***
	(135.43)	(135.49)	(135.44)	(135.35)	(135.39)	(135.38)	(135.33)	(134.61)
HF_{t-1}	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***
	(-4.98)	(-5.04)	(-5.01)	(-5.02)	(-4.99)	(-5.03)	(-5.02)	(-5.16)
$Subprime_{t-12}$	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(-7.14)	(-7.27)	(-7.16)	(-7.09)	(-7.14)	(-7.14)	(-7.22)	(-7.15)
$\Delta Loan\ supply_{t-12}$	0.0001	-0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
	(0.33)	(-0.04)	(0.47)	(0.37)	(0.30)	(0.31)	(0.28)	(0.70)
Recourse	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(0.11)	(0.11)	(0.13)	(0.12)	(0.10)	(0.14)	(0.14)	(0.30)
Constant	-0.0008***	-0.0008***	-0.0008***	-0.0008***	-0.0008***	-0.0008***	-0.0008***	-0.0009***
	(-3.29)	(-3.29)	(-3.36)	(-3.29)	(-3.27)	(-3.34)	(-3.38)	(-3.71)
Number of Obs	9,370	9,370	9,370	9,370	9,370	9,370	9,370	9,370
Adjusted R-squared	0.7036	0.7038	0.7039	0.7038	0.7036	0.7039	0.7039	0.7047

(Continued on next page)

(continued)

Panel B. 2013-2016

	ΔHP							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta \mathrm{MDRI}_{t-1}$	-0.0002						-0.0002	-0.0002
	(-0.43)						(-0.55)	(-0.59)
$\Delta \mathrm{MDRI}_{t-2}$		-0.0002					-0.0003	-0.0004
		(-0.56)					(-0.64)	(-0.85)
$\Delta \mathrm{MDRI}_{t-3}$			0.0001				0.0000	-0.0002
			(0.25)				(0.05)	(-0.45)
$\Delta \mathrm{MDRI}_{t-4}$				-0.0003				-0.0005
				(-0.82)				(-1.20)
$\Delta \mathrm{MDRI}_{t-5}$					-0.0006			-0.0006
					(-1.40)			(-1.53)
$\Delta \mathrm{MDRI}_{t-6}$						0.0003		0.0001
						(0.79)		(0.28)
$\Delta \mathrm{HP}_{t-1}$	0.8250***	0.8249***	0.8251***	0.8249***	0.8249***	0.8251***	0.8248***	0.8244***
	(147.90)	(147.90)	(147.93)	(147.93)	(147.91)	(147.94)	(147.69)	(147.34)
HF_{t-1}	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
	(-0.98)	(-0.98)	(-0.98)	(-1.00)	(-1.02)	(-0.97)	(-0.97)	(-1.03)
$\mathrm{Subprime}_{t-12}$	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(-4.05)	(-4.06)	(-4.05)	(-4.05)	(-4.04)	(-4.06)	(-4.06)	(-4.06)
$\Delta \mathrm{Loan\ supply}_{t-12}$	0.0007***	0.0007***	0.0007***	0.0007***	0.0007***	0.0007***	0.0007***	0.0006***
	(6.33)	(6.29)	(6.35)	(6.27)	(6.22)	(6.38)	(6.22)	(5.85)
Recourse	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***
	(-6.96)	(-6.95)	(-6.98)	(-6.95)	(-6.94)	(-6.99)	(-6.92)	(-6.83)
Constant	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002
	(-0.54)	(-0.55)	(-0.53)	(-0.54)	(-0.55)	(-0.53)	(-0.56)	(-0.61)
Number of Obs	9,884	9,884	9,884	9,884	9,884	9,884	9,884	9,884
Adjusted R-squared	0.7358	0.7358	0.7358	0.7358	0.7359	0.7358	0.7358	0.7358

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

Table B3: Predictive power of MDRI for fundamental house price appreciation rates.

	Fundamental house price appreciation rate							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta \mathrm{MDRI}_{t-1}$	-0.0007**						-0.0007**	-0.0007**
	(-2.46)						(-2.44)	(-2.31)
$\Delta \mathrm{MDRI}_{t-2}$		0.0001					-0.0001	-0.0003
		(0.21)					(-0.46)	(-0.96)
$\Delta \mathrm{MDRI}_{t-3}$			0.0001				0.0000	-0.0005*
			(0.46)				(0.00)	(-1.77)
$\Delta \mathrm{MDRI}_{t-4}$				-0.0007**				-0.0013***
				(-2.52)				(-4.13)
$\Delta \mathrm{MDRI}_{t-5}$					-0.0011***			-0.0016***
					(-4.17)			(-5.23)
$\Delta \mathrm{MDRI}_{t-6}$						0.0002		-0.0003
						(0.81)		(-1.13)
$\Delta \mathrm{Fundamental}_{t-1}$	0.8358***	0.8359***	0.8359***	0.8356***	0.8355***	0.8360***	0.8358***	0.8345***
	(217.27)	(217.24)	(217.20)	(217.16)	(217.19)	(217.13)	(217.14)	(216.35)
HF_{t-1}	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(-8.13)	(-8.16)	(-8.15)	(-8.20)	(-8.20)	(-8.14)	(-8.13)	(-8.30)
$\mathrm{Subprime}_{t-12}$	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***
	(-12.47)	(-12.45)	(-12.45)	(-12.42)	(-12.49)	(-12.45)	(-12.45)	(-12.43)
$\Delta \mathrm{Loan\ supply}_{t-12}$	0.0007***	0.0007***	0.0007***	0.0007***	0.0007***	0.0007***	0.0007***	0.0007***
	(7.69)	(7.68)	(7.69)	(7.63)	(7.56)	(7.71)	(7.71)	(7.40)
Recourse	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(-3.72)	(-3.78)	(-3.79)	(-3.73)	(-3.69)	(-3.80)	(-3.69)	(-3.40)
Constant	-0.0006***	-0.0006***	-0.0006***	-0.0006***	-0.0006***	-0.0006***	-0.0006***	-0.0007***
	(-4.38)	(-4.27)	(-4.25)	(-4.31)	(-4.43)	(-4.25)	(-4.39)	(-4.77)
Number of Obs	19,254	19,254	19,254	19,254	19,254	19,254	19,254	19,254
Adjusted R-squared	0.7850	0.7849	0.7849	0.7850	0.7851	0.7849	0.7850	0.7853

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

Table B4: Predictive power of MDRI for the bubble component of house prices.

	Change in bubble component							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ΔMDRI_{t-1}	0.0001						0.0001	0.0001
	(0.32)						(0.26)	(0.25)
ΔMDRI_{t-2}		0.0004					0.0002	0.0002
		(1.28)					(0.72)	(0.49)
ΔMDRI_{t-3}			-0.0007**				-0.0006**	-0.0007**
			(-2.35)				(-1.98)	(-2.06)
ΔMDRI_{t-4}				-0.0003				-0.0005
				(-0.95)				(-1.52)
ΔMDRI_{t-5}					0.0005*			0.0001
					(1.87)			(0.31)
ΔMDRI_{t-6}						-0.0008***		-0.0008**
						(-2.67)		(-2.44)
$\Delta\text{Bubble}_{t-1}$	0.3056***	0.3056***	0.3057***	0.3056***	0.3057***	0.3058***	0.3058***	0.3059***
	(44.88)	(44.88)	(44.90)	(44.87)	(44.90)	(44.91)	(44.90)	(44.93)
HF_{t-1}	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***
	(-3.34)	(-3.35)	(-3.34)	(-3.34)	(-3.32)	(-3.35)	(-3.36)	(-3.39)
Subprime_{t-12}	-0.0001**	-0.0001**	-0.0001**	-0.0001**	-0.0001**	-0.0001**	-0.0001**	-0.0001**
	(-2.46)	(-2.51)	(-2.46)	(-2.44)	(-2.46)	(-2.42)	(-2.49)	(-2.41)
$\Delta\text{Loan supply}_{t-12}$	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
	(-0.35)	(-0.40)	(-0.38)	(-0.38)	(-0.28)	(-0.43)	(-0.40)	(-0.51)
Recourse	-0.0001**	-0.0001**	-0.0001**	-0.0001**	-0.0001**	-0.0001**	-0.0001**	-0.0001**
	(-2.29)	(-2.32)	(-2.23)	(-2.27)	(-2.33)	(-2.22)	(-2.26)	(-2.16)
Constant	-0.0000	-0.0000	-0.0000	-0.0000	0.0000	-0.0000	-0.0000	-0.0000
	(-0.03)	(-0.03)	(-0.13)	(-0.05)	(0.02)	(-0.10)	(-0.10)	(-0.17)
Number of Obs	19,254	19,254	19,254	19,254	19,254	19,254	19,254	19,254
Adjusted R-squared	0.0966	0.0967	0.0969	0.0967	0.0968	0.0970	0.0968	0.0971

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

Table B5: Predictive power of MDRI for the Homes Foreclosed.

Panel A. 2007-2012

	lhf							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ΔMDRI_{t-1}	0.0071						-0.0607***	-0.0625***
	(0.42)						(-3.25)	(-3.33)
ΔMDRI_{t-2}		-0.1048***					-0.1633***	-0.1638***
		(-6.10)					(-8.17)	(-7.89)
ΔMDRI_{t-3}			-0.0453***				-0.1080***	-0.0963***
			(-2.68)				(-5.79)	(-4.67)
ΔMDRI_{t-4}				0.0356**				0.0055
				(2.15)				(0.28)
ΔMDRI_{t-5}					0.0508***			0.0516***
					(3.08)			(2.66)
ΔMDRI_{t-6}						-0.0607***		-0.0258
						(-3.69)		(-1.44)
HF_{t-1}	0.9729***	0.9731***	0.9729***	0.9729***	0.9730***	0.9728***	0.9733***	0.9734***
	(435.99)	(436.92)	(436.18)	(436.11)	(436.24)	(436.26)	(437.74)	(437.86)
ΔHP_{t-1}	-0.7085***	-0.7419***	-0.7191***	-0.6954***	-0.6907***	-0.7315***	-0.7938***	-0.7794***
	(-3.29)	(-3.45)	(-3.34)	(-3.23)	(-3.21)	(-3.40)	(-3.69)	(-3.62)
Subprime_{t-12}	0.0055***	0.0060***	0.0055***	0.0054***	0.0055***	0.0055***	0.0062***	0.0062***
	(4.74)	(5.16)	(4.72)	(4.70)	(4.80)	(4.75)	(5.32)	(5.38)
$\Delta \text{Loan supply}_{t-12}$	-0.0136*	-0.0068	-0.0123*	-0.0139*	-0.0140**	-0.0134*	0.0011	0.0003
	(-1.91)	(-0.94)	(-1.73)	(-1.95)	(-1.96)	(-1.89)	(0.15)	(0.04)
Recourse	-0.0193***	-0.0194***	-0.0191***	-0.0193***	-0.0194***	-0.0190***	-0.0190***	-0.0190***
	(-4.63)	(-4.67)	(-4.59)	(-4.64)	(-4.66)	(-4.58)	(-4.57)	(-4.59)
Constant	0.0811***	0.0816***	0.0801***	0.0811***	0.0819***	0.0802***	0.0787***	0.0795***
	(9.60)	(9.69)	(9.49)	(9.61)	(9.70)	(9.51)	(9.34)	(9.42)
Number of Obs	9,316	9,316	9,316	9,316	9,316	9,316	9,316	9,316
Adjusted R-squared	0.9644	0.9646	0.9645	0.9645	0.9645	0.9645	0.9647	0.9648

(Continued on next page)

(continued)
Panel B. 2013-2016

	lhf							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ΔMDRI_{t-1}	0.0750**						0.0168	0.0015
	(2.21)						(0.48)	(0.04)
ΔMDRI_{t-2}		-0.1787***					-0.1991***	-0.2257***
		(-5.19)					(-5.50)	(-6.14)
ΔMDRI_{t-3}			-0.0914***				-0.1278***	-0.1780***
			(-2.62)				(-3.55)	(-4.74)
ΔMDRI_{t-4}				-0.0180				-0.1458***
				(-0.51)				(-3.82)
ΔMDRI_{t-5}					-0.0963***			-0.1791***
					(-2.72)			(-4.77)
ΔMDRI_{t-6}						-0.1491***		-0.1941***
						(-4.23)		(-5.32)
HF_{t-1}	0.9638***	0.9640***	0.9637***	0.9637***	0.9636***	0.9635***	0.9639***	0.9629***
	(293.79)	(294.16)	(293.76)	(293.66)	(293.70)	(293.86)	(294.28)	(294.34)
ΔHP_{t-1}	-0.2070	-0.3046	-0.2683	-0.2411	-0.2667	-0.2929	-0.3502	-0.5508
	(-0.42)	(-0.61)	(-0.54)	(-0.48)	(-0.54)	(-0.59)	(-0.70)	(-1.11)
Subprime_{t-12}	0.0033	0.0031	0.0033	0.0033	0.0034	0.0037	0.0030	0.0035
	(0.67)	(0.62)	(0.67)	(0.67)	(0.69)	(0.74)	(0.60)	(0.70)
$\Delta\text{Loan supply}_{t-12}$	-0.0644***	-0.0689***	-0.0670***	-0.0657***	-0.0674***	-0.0671***	-0.0714***	-0.0824***
	(-6.80)	(-7.27)	(-7.07)	(-6.92)	(-7.10)	(-7.09)	(-7.50)	(-8.52)
Recourse	0.0027	0.0045	0.0034	0.0032	0.0035	0.0037	0.0049	0.0076
	(0.41)	(0.68)	(0.50)	(0.48)	(0.53)	(0.56)	(0.74)	(1.15)
Constant	0.0336	0.0273	0.0311	0.0322	0.0316	0.0322	0.0250	0.0192
	(1.06)	(0.86)	(0.98)	(1.01)	(1.00)	(1.02)	(0.79)	(0.61)
Number of Obs	9,832	9,832	9,832	9,832	9,832	9,832	9,832	9,832
Adjusted R-squared	0.8998	0.9000	0.8998	0.8997	0.8998	0.8999	0.9001	0.9006

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

Appendix C

Table C1: The correlation coefficients for U.S. country-level ASVI and other variables.

	USASVI1	*USASVI2*	*USASVI3*	*USASVI4*	*USASVI5*	*USASVI6*	*USASVI7*	*USASVI8*	*USASVI9*
USASVI1	1***								
USASVI2	0.839***	1***							
USASVI3	0.623***	0.91***	1***						
USASVI4	0.718***	0.671***	0.618***	1***					
USASVI5	0.218***	0.326***	0.445***	0.268***	1***				
USASVI6	0.151***	0.38***	0.474***	0.177***	0.261***	1***			
USASVI7	0.447***	0.811***	0.889***	0.411***	0.43***	0.576***	1***		
USASVI8	0.385***	0.756***	0.837***	0.382***	0.475***	0.602***	0.983***	1***	
USASVI9	0.501***	0.799***	0.731***	0.483***	0.064***	0.223***	0.698***	0.656***	1***
USASVI10	0.408***	0.718***	0.815***	0.482***	0.596***	0.55***	0.905***	0.926***	0.556***
USASVI11	0.801***	0.913***	0.892***	0.775***	0.625***	0.378***	0.76***	0.732***	0.629***
USASVI12	0.496***	0.863***	0.881***	0.476***	0.297***	0.495***	0.942***	0.921***	0.892***
ΔDELQ	0.047**	0.136***	0.217***	-0.002	0.53***	0.39***	0.369***	0.408***	-0.04**
FS	-0.094***	-0.016	-0.006	0.032	-0.005	0.255***	0.07***	0.093***	0.07***
ΔHP	-0.144***	-0.209***	-0.202***	-0.241***	-0.12***	-0.329***	-0.237***	-0.245***	-0.172***
ΔIncome	-0.157***	-0.325***	-0.395***	-0.121***	-0.273***	-0.369***	-0.574***	-0.583***	-0.228***
ΔUnemp	0.41***	0.526***	0.517***	0.37***	0.348***	0.487***	0.63***	0.642***	0.343***
Highschool_pct	*−0.07****	*−0.056****	*−0.046***	*−0.123****	*−0.056****	*−0.136****	*−0.053****	*−0.063****	*−0.038**

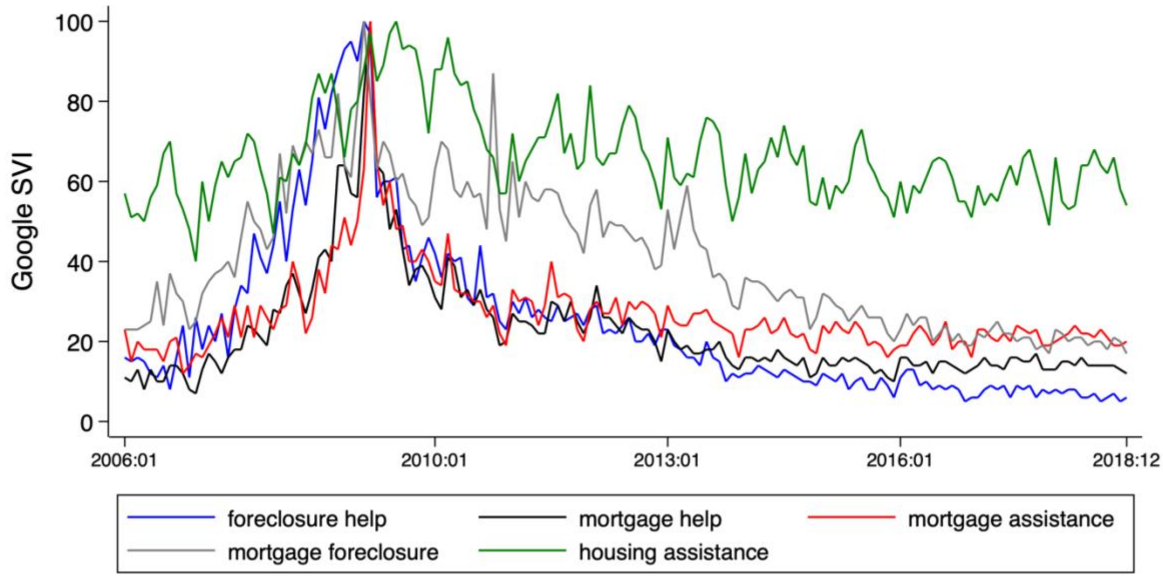
(Continued on next page)

(continued)

	USASVI10	***USASVI11***	***USASVI12***	***ΔDELQ***	***FS***	***ΔHP***	***ΔIncome***	***ΔUnemp***	***Highschol_pct***
USASVI10	1***								
USASVI11	0.776***	1***							
USASVI12	0.831***	0.755***	1***						
ΔDELQ	0.501***	0.277***	0.228***	1***					
FS	0.135***	-0.005	0.09***	0.161***	1***				
ΔHP	-0.342***	-0.24***	-0.238***	-0.357***	-0.686***	1***			
ΔIncome	-0.512***	-0.337***	-0.466***	-0.391***	-0.246***	0.276***	1***		
ΔUnemp	0.661***	0.556***	0.564***	0.508***	0.262***	-0.467***	-0.555***	1***	
Highschool_pct	*–0.116****	*–0.089****	*–0.06****	*–0.101****	*–0.382****	*0.232****	*0.092****	*–0.128****	*1****

Notes: This table presents the correlation coefficients for the country-level abnormal SVI index and other main variables used in the empirical sections. *, **, and *** denote that the coefficient estimates are significant at 10%, 5%, and 1% statistical levels, respectively.

Panel A. Google SVI for independent search terms in the naïve search group



Panel B. Google SVI for independent search terms in the sophisticated search group

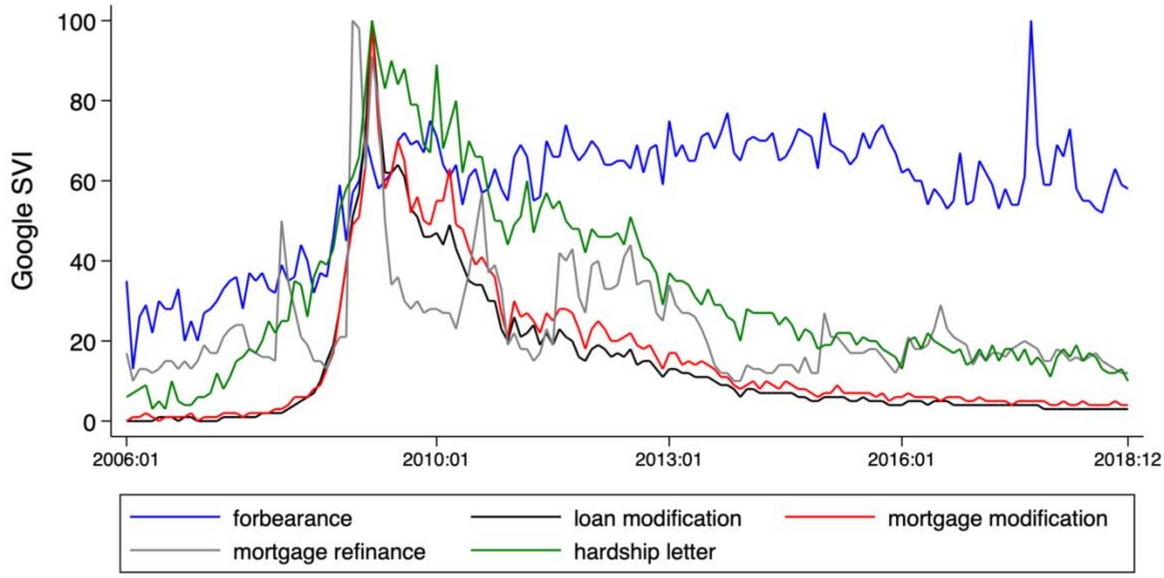


Figure C1: Dynamics of the Google SVI for independent search terms.

Notes: The search terms for each line are given by the label at the bottom of the figure.

References

Abraham, J. M., & Hendershott, P. H. (1996). Bubbles in metropolitan housing markets. *Journal of Housing Research*, 7(2), 191-207.

Abrigo, M. R., & Love, I. (2016). Estimation of panel vector autoregression in Stata. *Stata Journal*, 16(3), 778-804.

Afkhami, M., Cormack, L., & Ghoddusi, H. (2017). Google search keywords that best predict energy price volatility. *Energy Economics*, 67, 17-27.

Agarwal, S., Chomsisengphet, S., & Zhang, Y. (2017). How does working in a finance profession affect mortgage delinquency? *Journal of Banking and Finance*, 78, 1-13.

Ambrose, B. W., Buttimer Jr, R. J., & Capone, C. A. (1997). Pricing mortgage default and foreclosure delay. *Journal of Money, Credit, and Banking*, 29(3), 314-325.

Amromin, G., & Paulson, A. L. (2009). Comparing patterns of default among prime and subprime mortgages. *Economic Perspectives*, 33(2).

Anderson, J. T., Freybote, J., Lucus, D., Seiler, M. J., & Simon, L. (2022). Using Artificial Intelligence to Identify Strategic Mortgage Default Attitudes. *Journal of Real Estate Research*, 44(3), 429-445.

Anenberg, E., & Kung, E. (2014). Estimates of the size and source of price declines due to nearby foreclosures. *American Economic Review*, 104(8), 2527-2551.

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1), 29-51.

Arguello, J., Wu, W. C., Kelly, D., & Edwards, A. (2012). Task complexity, vertical display and user interaction in aggregated search. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 435-444.

Arnio, A. N., Baumer, E. P., & Wolff, K. T. (2012). The contemporary foreclosure crisis and US crime rates. *Social Science Research*, 41(6), 1598-1614.

Aroul, R. R., Sabherwal, S., & Saydometov, S. (2022). FEAR Index, city characteristics, and housing returns. *Real Estate Economics*, 50(1), 173-205.

Ashworth, J., & Parker, S. (1997). Modelling regional house prices in the UK. *Scottish Journal of Political Economy*, 44(3), 225-246.

Askitas, N., & Zimmermann, K. F. (2011). Detecting mortgage delinquencies. IZA Discussion Paper No. 5895. Available at SSRN: https://ssrn.com/abstract=1906198 (Accessed: 31 August 2022)

Avery, R. B., Brevoort, K. P., & Canner, G. B. (2007). The 2006 HMDA data. *Federal Reserve Bulletin*, 93, A73-A109.

Bajari, P., Chu, C. S., & Park, M. (2008). An empirical model of subprime mortgage default from 2000 to 2007. National Bureau of Economic Research Working Paper No. w14625, Available at: https://www.nber.org/papers/w14625 (Accessed: 31 August 2022)

Baker, S. R., & Fradkin, A. (2017). The impact of unemployment insurance on job search: Evidence from Google search data. *Review of Economics and Statistics*, 99(5), 756-768.

Beracha, E., & Wintoki, M. B. (2013). Forecasting residential real estate price changes from online search activity. *Journal of Real Estate Research*, 35(3), 283-312.

Bhutta, N., Dokko, J., & Shan, H. (2017). Consumer ruthlessness and mortgage default during the 2007 to 2009 housing bust. *Journal of Finance*, 72(6), 2433-2466.

Biswas, A., Fout, H., & Pennington-Cross, A. (2020). Mortgage losses under alternative property disposition approaches: Deed-in-Lieu, short sales, and foreclosure sales. *Journal of Real Estate Finance and Economics*, 1-33.

Boor, C.D. (1978). *A Practical Guide to Splines*. New York, NY: Springer.

Brand-Gruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving by experts and novices: Analysis of a complex cognitive skill. *Computers in Human Behaviour*, 21(3), 487-508.

Calomiris, C. W., Longhofer, S. D., & Miles, W. R. (2013). The foreclosure–house price nexus: A Panel VAR model for U.S. states, 1981–2009. *Real Estate Economics*, 41(4), 709-746.

Campbell, J. Y., Giglio, S., & Pathak, P. (2011). Forced sales and house prices. *American Economic Review*, 101(5), 2108-2131.

Canova, F., & Ciccarelli, M. (2013). Panel vector autoregressive models: A survey. European Central Bank Working Paper No. 1507. Available at: http://dx.doi.org/10.2139/ssrn.2201610 (Accessed: 31 August 2022)

Capozza, D. R., Hendershott, P. H., & Mack, C. (2004). An anatomy of price dynamics in illiquid markets: Analysis and evidence from local housing markets. *Real Estate Economics*, 32(1), 1-32.

Capozza, D. R., & Seguin, P. J. (1996). Expectations, efficiency, and euphoria in the housing market. *Regional Science and Urban Economics*, 26(3-4), 369-386.

Case, K. E., & Shiller, R. J. (1989). The efficiency of the market for single-family homes. *American Economic Review*, 79(1), 125-137.

Case, K. E., & Shiller, R. J. (1990). Forecasting prices and excess returns in the housing market. *Real Estate Economics*, 18(3), pp.253-273.

Chauvet, M., Gabriel, S., & Lutz, C. (2016). Mortgage default risk: New evidence from internet search queries. *Journal of Urban Economics*, 96, 91-111.

Choi, I. (2001). Unit root tests for panel data. *Journal of International Money and Finance*, 20(2), 249-272.

Clauretie, T. M., & Daneshvary, N. (2011). The optimal choice for lenders facing defaults: Short sale, foreclose, or REO. *Journal of Real Estate Finance and Economics*, 42(4), 504-521.

Cui, L., & Walsh, R. (2015). Foreclosure, vacancy and crime. *Journal of Urban Economics*, 87, 72-84.

Cunningham, C., Gerardi, K., & Shen, L. (2021). The double trigger for mortgage default: Evidence from the fracking boom. *Management Science*, 67(6), 3943-3964.

Cutts, A. C., and W. Merrill. (2008). Interventions in mortgage default: Policies and practices to prevent home loss and lower costs. Freddie Mac Working Paper No.08-01.

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *Journal of Finance*, 66(5), 1461-1499.

Da, Z., Engelberg, J., & Gao, P. (2015). The sum of all FEARS investor sentiment and asset prices. *Review of Financial Studies*, 28(1), 1-32.

Damianov, D. S., Wang, X., & Yan, C. (2021). Google search queries, foreclosures, and house prices. *Journal of Real Estate Finance and Economics*, 63(2), 177-209.

Daneshvary, N., Clauretie, T., & Kader, A. (2011). Short-term own-price and spillover effects of distressed residential properties: The case of a housing crash. *Journal of Real Estate Research*, 33(2), 179-208.

Daneshvary, N., & Clauretie, T. M. (2012). Toxic neighbors: Foreclosures and short-sales spillover effects from the current housing-market crash. *Economic Inquiry*, 50(1), 217-231.

Demiroglu, C., Dudley, E., & James, C. M. (2014). State foreclosure laws and the incidence of mortgage default. *Journal of Law and Economics*, 57(1), 225-280.

Deng, Y., Quigley, J. M., & Van Order, R. (2000). Mortgage terminations, heterogeneity and the exercise of mortgage options. *Econometrica*, 68(2), 275-307.

Depken, C. A., Hollans, H., & Swidler, S. (2015). A low-cost methodology for correcting the distressed sales bias in a downward spiraling housing market. *Journal of Real Estate Research*, 37(1), 151-172.

Dorsey, R. E., Hu, H., Mayer, W. J., & Wang, H. (2010). Hedonic versus repeat-lases house price indexes for measuring the recent boom-bust cycle. *Journal of Housing Economics*, 19(2), 75-93.

Drake, L. (1993). Modelling UK house prices using cointegration: An application of the Johansen technique. *Applied Economics*, 25(9), 1225-1228.

Eichengreen, B., Mody, A., Nedeljkovic, M., & Sarno, L. (2012). How the subprime crisis went global: Evidence from bank credit default swap spreads. *Journal of International Money and Finance*, 31(5), 1299-1318.

Eickhoff, C., Teevan, J., White, R., & Dumais, S. (2014). Lessons from the journey: A query log analysis of within-session learning. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, 223-232.

Ellen, I. G., Lacoe, J., & Sharygin, C. A. (2013). Do foreclosures cause crime? *Journal of Urban Economics*, 74, 59-70.

Elmer, P., & Seelig, S. (1999). Insolvency, trigger events, and consumer risk posture in the theory of single-family mortgage default. *Journal of Housing Research*, 10(1), 1-25.

Elul, R., Souleles, N. S., Chomsisengphet, S., Glennon, D., & Hunt, R. M. (2010). What 'triggers' mortgage default? *American Economic Review*, 100(2), 490-94.

Favara, G., & Imbs, J. (2015). Credit supply and the price of housing. *American Economic Review*, 105(3), 958-92.

Foote, C. L., Gerardi, K., & Willen, P. S. (2008). Negative equity and foreclosure: Theory and evidence. *Journal of Urban Economics*, 64(2), 234-245.

Foster, C., & Van Order, R. (1984). An option-based model of mortgage default. *Housing Finance Review*, 3(4), 351.

Gadiraju, U., Yu, R., Dietze, S., & Holtz, P. (2018). Analyzing knowledge gain of users in informational search sessions on the web. In *Proceedings of the 2018 Conference on Human Information Interaction and Retrieval*, 2-11.

Gao, Z., Ren, H., & Zhang, B. (2020). Googling investor sentiment around the world. *Journal of Financial and Quantitative Analysis*, 55(2), 549-580.

Gerardi, K., Herkenhoff, K. F., Ohanian, L. E., & Willen, P. S. (2018). Can't pay or won't pay? Unemployment, negative equity, and strategic default. *Review of Financial Studies*, 31(3), 1098-1131.

Gerardi, K., Rosenblatt, E., Willen, P. S., & Yao, V. (2015). Foreclosure externalities: New evidence. *Journal of Urban Economics*, 87, 42-56.

Gerardi, K., Shapiro, A. H., & Willen, P. (2007). Subprime outcomes: Risky mortgages, homeownership experiences, and foreclosures. Federal Reserve Bank of Boston Working Paper No. 07-15. Available at https://ssrn.com/abstract=1073182 (Accessed: 31 August 2022)

Ghent, A. C., & Kudlyak, M. (2011). Recourse and residential mortgage default: Evidence from U.S. states. *Review of Financial Studies*, 24(9), 3139-3186.

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

Goodwin, K. R., & Johnson, K. H. (2017). The short sale stigma. *Journal of Real Estate Finance and Economics*, 55(4), 416-434.

Guiso, L., Sapienza, P., & Zingales, L. (2013). The determinants of attitudes toward strategic default on mortgages. *Journal of Finance*, 68(4), 1473-1515.

Harding, J. P., Rosenblatt, E., & Yao, V. W. (2009). The contagion effect of foreclosed properties. *Journal of Urban Economics*, 66(3), 164-178.

Hartley, D. (2014). The effect of foreclosures on nearby housing prices: Supply or dis-amenity? *Regional Science and Urban Economics*, 49, 108-117.

Hastings, J. S., & Tejeda-Ashton, L. (2008). Financial literacy, information, and demand elasticity: Survey and experimental evidence from Mexico. National Bureau of Economic Research Working Paper No. w14538. Available at https://www.nber.org/papers/w14538 (Accessed: 09 October 2023)

Hersh, W. R., Crabtree, M. K., Hickam, D. H., Sacherek, L., Friedman, C. P., Tidmarsh, P., Mosbaek, C., & Kraemer, D. (2002). Factors associated with success in searching

MEDLINE and applying evidence to answer clinical questions. *Journal of the American Medical Informatics Association*, 9(3), 283-293.

Hersh, W. R., Elliot, D. L., Hickam, D. H., Wolf, S. L., & Molnar, A. (1995). Towards new measures of information retrieval evaluation. In *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 164-170.

Holtz-Eakin, D., Newey, W., & Rosen, H. (1988). Estimating vector autoregressions with panel data. *Econometrica*, 56(6), 1371-1395.

Houle, J. N. (2014). Mental health in the foreclosure crisis. *Social Science and Medicine*, 118, 1-8.

Illies, J.J., & Reiter-Palmon, R. (2004). The effects of type and level of personal involvement on information search and problem solving. *Journal of Applied Social Psychology*, 34(8), 1709-1729.

Judson, R. A., & Owen, A. L. (1999). Estimating dynamic panel data models: A guide for macroeconomists. *Economics Letters*, 65(1), 9-15.

Kasparova, D, & White, M. (2001). The responsiveness of house prices to macroeconomic forces: A cross country comparison. *European Journal of Housing Policy*, 1(3), 385-416.

Kau, J. B., Keenan, D. C., & Kim, T. (1994). Default probabilities for mortgages. *Journal of Urban Economics*, 35(3), 278-296.

Kelly, R., & McCann, F. (2016). Some defaults are deeper than others: Understanding long-term mortgage arrears. *Journal of Banking and Finance*, 72, 15-27.

Lei, P. L., Lin, S. S., & Sun, C. T. (2013). Effect of reading ability and Internet experience on keyword-based image search. *Journal of Educational Technology and Society*, 16(2), 151-162.

Li, X., Ma, J., Wang, S., & Zhang, X. (2015). How does Google search affect trader positions and crude oil prices? *Economic Modelling*, 49, 162-171.

Libman, K., Fields, D., & Saegert, S. (2012). Housing and health: A social ecological perspective on the US foreclosure crisis. *Housing, Theory and Society*, 29(1), 1-24.

Lin, Z., Rosenblatt, E., & Yao, V. W. (2009). Spillover effects of foreclosures on neighborhood property values. *Journal of Real Estate Finance and Economics*, 38(4), 387-407.

Love, I., & Zicchino, L. (2006). Financial development and dynamic investment behaviour: Evidence from panel VAR. *Quarterly Review of Economics and Finance*, 46(2), 190-210.

Lusardi, A., & Mitchelli, O.S. (2007). Financial literacy and retirement preparedness: Evidence and implications for financial education: The problems are serious, and remedies are not simple. *Business Economics*, 42, 35-44.

Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *American Economic Journal: Journal of Economic Literature*, 52(1), 5-44.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Mayer, C. (2011). Housing bubbles: A survey. *Annual Review of Economics*, 3(1), 559-577.

Mayer, C., & Pence, K. (2008). Subprime mortgages: What, where, and to whom? National Bureau of Economic Research Working Paper No. 14083. Available at https://www.nber.org/papers/w14083 (Accessed: 31 August 2022)

McLaren, N., & Shanbhogue, R. (2011). Using internet search data as economic indicators. Bank of England Quarterly Bulletin No. 2011 Q2. Available at http://dx.doi.org/10.2139/ssrn.1865276 (Accessed: 31 August 2022)

Mendoza, E. G., & Quadrini, V. (2010). Financial globalization, financial crises and contagion. *Journal of Monetary Economics*, 57(1), 24-39.

Mian, A., Sufi, A., & Trebbi, F. (2015). Foreclosures, house prices, and the real economy. *Journal of Finance*, 70(6), 2587-2634.

Mocetti, S., & Viviano, E. (2017). Looking behind mortgage delinquencies. *Journal of Banking and Finance*, 75, 53-63.

Monchaux, S., Amadieu, F., Chevalier, A., & Mariné, C. (2015). Query strategies during information searching: Effects of prior domain knowledge and complexity of the information problems to be solved. *Information Processing and Management*, 51(5), 557-569.

Nordlie, R. (1999). "User revealment"—A comparison of initial queries and ensuing question development in online searching and in human reference interactions. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, 11-18.

Orgaz-Molina, J., Cotugno, M., Girón-Prieto, M. S., Arrabal-Polo, M. A., Ruiz-Carrascosa, J. C., Buendía-Eisman, A., & Arias-Santiago, S. (2015). A study of internet searches for medical information in dermatology patients: The patient–physician relationship. *Actas dermo-sifiliograficas*, 106(6), 493-499.

Pesaran, H. H., & Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. *Economics letters*, 58(1), 17-29.

Rogers, W., & Winter, W. (2009). The impact of foreclosures on neighboring housing sales. *Journal of Real Estate Research*, 31(4), 455-480.

Roscoe, R. D., Grebitus, C., O'Brian, J., Johnson, A. C., & Kula, I. (2016). Online information search and decision making: Effects of web search stance. *Computers in Human Behaviour*, 56, 103-118.

Rose, L. A. (1989). Urban land supply: Natural and contrived restrictions. *Journal of Urban Economics*, 25(3), 325-345.

Saiz, A. (2010). The geographic determinants of housing supply. *Quarterly Journal of Economics*, *125*(3), 1253-1296.

Sanchiz, M., Chevalier, A., & Amadieu, F. (2017a). How do older and young adults start searching for information? Impact of age, domain knowledge and problem complexity on the different steps of information searching. *Computers in Human Behaviour*, 72, 67-78.

Sanchiz, M., Chin, J., Chevalier, A., Fu, W. T., Amadieu, F., & He, J. (2017b). Searching for information on the Web: Impact of cognitive aging, prior domain knowledge and complexity of the search problems. *Information Processing and Management*, 53(1), 281-294.

Seiler, M. J. (2016). The perceived moral reprehensibility of strategic mortgage default. *Journal of Housing Economics*, 32, 18-28.

Singer, E. & Ye, C. (2013). The use and effects of incentives in surveys. *Annals of the American Academy of Political and Social Science*, 645(1), 112-141.

Stevenson, S. (2008). Modeling housing market fundamentals: Empirical evidence of extreme conditions. *Real Estate Economics*, 36(1), 1-29.

Sullivan, D, 2004, Google tops, but Yahoo switch success so far. Available at https://www.searchenginewatch.com/2004/04/04/google-tops-but-yahoo-switch-success-so-far/ (Accessed: 31 August 2022)

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

Tian, C. Y., Quercia, R. G., & Riley, S. (2016). Unemployment as an adverse trigger event for mortgage default. *Journal of Real Estate Finance and Economics*, 52(1), 28-49.

Vakkari, P., Pennanen, M., & Serola, S. (2003). Changes of search terms and tactics while writing a research proposal: A longitudinal case study. *Information Processing and Management*, 39(3), 445-463.

Van Dijk, D. W., & Francke, M. K. (2018). Internet search behaviour, liquidity and prices in the housing market. *Real Estate Economics*, 46(2), 368-403.

Visser, E. B., & Weideman, M. (2011). An empirical study on website usability elements and how they affect search engine optimisation. *South African Journal of Information Management*, 13(1), 1-9.

Walhout, J., Oomen, P., Jarodzka, H., & Brand-Gruwel, S. (2017). Effects of task complexity on online search behaviour of adolescents. *Journal of the Association for Information Science and Technology*, 68(6), 1449-1461.

Webb, G. K. (2009). Internet search statistics as a source of business intelligence: Searches on foreclosure as an estimate of actual home foreclosures. *Issues in Information Systems*, 10(2), 82.

Wildemuth, B. M. (2004). The effects of domain knowledge on search tactic formulation. *Journal of the American Society for Information Science and Technology*, 55(3), 246-258.

Wirth, W., Sommer, K., Von Pape, T., & Karnowski, V. (2016). Success in online searches: Differences between evaluation and finding tasks. *Journal of the Association for Information Science and Technology*, 67(12), 2897-2908.

Wu, L., & Brynjolfsson, E. (2015). The future of prediction: How Google searches foreshadow housing prices and sales. *Economic analysis of the digital economy*, 89-118.

Yan, C., Phylaktis, K., & Fuertes, A. M. (2016). On cross-border bank credit and the U.S. financial crisis transmission to FX markets. *Journal of International Money and Finance*, 69, 108-134.

Yu, L., Zhao, Y., Tang, L., & Yang, Z. (2019). Online big data-driven oil consumption forecasting with Google trends. *International Journal of Forecasting*, 35(1), 213-223.

Ziebland, S., Chapple, A., Dumelow, C., Evans, J., Prinjha, S., & Rozmovits, L. (2004). How the internet affects patients' experience of cancer: A qualitative study. *BMJ*, *328*(7439), 564.

Zhu, S., & Pace, R. K. (2015). The influence of foreclosure delays on borrowers' default behaviour. *Journal of Money, Credit and Banking*, 47(6), 1205-1222.