Information disclosure vs. information learning via Google search

14 January 2024

Damian S. Damianov, Xiangdong Wang and Cheng Yan

Abstract

We decompose the Google Trends Search Volume Index into naïve and sophisticated searches and examine their impacts on mortgage default, respectively. Using U.S. data from 2006 to 2018, we find that the sophisticated search activity has a positive and robust relationship with the change in the percentage of mortgages in 90+ days of delinquency. However, foreclosure starts are positively related to naïve search activity in the short term, but negatively related to sophisticated search activity in the long term. Borrowers are more likely learn from sophisticated online searches than from naïve online searches, and they can use the information to avoid foreclosure starts and keep their houses. The relationship between Google search activity and mortgage default outcomes are significantly stronger in states that experienced substantial house price drops in the recent year. Our findings are robust to a battery of alternative settings.

Keywords: Mortgage default risk, Foreclosures, Google search
JEL classification:

Damianov: damian.damianov@durham.ac.uk, Durham University Business School, Millhill Lane, Durham, DH1 3LB, UK; Wang: xiangdong.wang@brunel.ac.uk, Department of Economics and Finance, Brunel University London, Kingston Ln, Uxbridge, London, UB8 3PH; Yan: Cheng.Yan.1@cass.city.ac.uk, Essex Business School, University of Essex, Colchester, CO4 3SQ, UK.
We acknowledge helpful comments received from Hugh Colaco, Philip Davis, Vasilis Sarafidis, and other participants at the BRG Internal Seminar at Brunel University London.

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, 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 41). 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 1]

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 stronger in states that experienced 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 2 reviews the literature on the current use of internet search data in finance and economics and discusses the hypothesis development. Section 3 introduces options provided by Fannie Mae to delinquent borrowers to keep their homes. Section 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 5, and Section 6 concludes.

2 Literature review

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.

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.

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 take action 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, we propose the following hypothesis:

Hypothesis 4*: Households use online searches to verify their pre-existing knowledge instead of learning from it.*

3 Options to stay at home

During the 2007 subprime crisis, one of the main concerns of the US 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 the borrower 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 induce 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 health and mental health of households (Houle, 2014; Libman et al., 2012). All of these outcomes encourage the borrower 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: [1]

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. 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

[1] 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

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 Data

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.[2] 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 behavior affects mortgage default, this study categorizes and contrasts two kinds of search activities: naïve and

[2] 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.

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 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 1.[3]

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 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 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 1.[4]

[Table 1]

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.

[3] 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.

[4] 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.

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 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. [5] 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. [6] 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 1.

Figure 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.

[Figure 2]

[5] 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.
[6] 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.

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.[7] The calculation is as follows:

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

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 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 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:

$$ASVIS_{i,t} = \sum_{j=0}^{6}\left(\theta\, \Delta Mtg30_{t-j}\right) + \delta_i + \varepsilon_{i,t} \quad (2)$$

where $ASVIS_{i,t}$ is the abnormal SVI reflecting the sophisticated search activity in U.S. state *i* at month *t* from Equation (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

[7] 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.

errors. Equation (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 $\widehat{ASVIS}_{i,t}$, and calculate the residual of Equation (2) as follows:

$$RASVIS_{i,t} = ASVIS_{i,t} - \widehat{ASVIS}_{i,t} \quad (3)$$

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 ($\widehat{ASVIS}_{i,t}$) 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 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 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 3]

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 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 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 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.

[Table 2 and Table 3]

5 Empirical results

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:

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

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 presents the estimation results for Equation (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.

[Table 4]

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.

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.[8] The new equation is as follows:

$$Default_{i,t} = \sum_{j=1,3} \left(\alpha_{1,j} ASVIN_{i,t-j} + \alpha_{2,j} RASVIS_{i,t-j} \right) + \alpha_3 \Delta Loansum_{i,t} + \alpha_4 \Delta Subprime_{i,t} + \sum_{m} \beta_m Controls_{i,t-m}^{m} + Year_t + \delta_i + \varepsilon_{i,t} \quad (5)$$

The estimation results for Equation (5) are presented in Table 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).

[Table 5]

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.

[8] 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.

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:

$$
\begin{aligned}
Default_{i,t} = & \sum_{j=1,3} \left(\alpha_{1,j} ASVIN_{i,t-j} + \alpha_{2,j} RASVIS_{i,t-j}\right) \\
& + \sum_{j=1,3} \left(\alpha_{3,j} SHPD_{i,t-j} + \alpha_{4,j} (SHPD \times ASVIN)_{i,t-j} + \alpha_{5,j} (SHPD \times RASVIS)_{i,t-j}\right) \\
& + \sum_{m} \beta_m Controls_{i,t-m}^{m} + Year_t + \delta_i + \varepsilon_{i,t} \qquad (6)
\end{aligned}
$$

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 (6) are presented in Table 6. The coefficients for ASVI, i.e., α_1 and $\alpha_2$in Equation (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 foreclosure.

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

[Table 6]

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.

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.[9] 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

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

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:

$$FS_{i,t} = \sum_{j=1,3} \left(\alpha_{1,j} ASVIN_{i,t-j} + \alpha_{2,j} RASVIS_{i,t-j}\right) \\ + \sum_{j=1,3} \left(\alpha_{3,j} \Delta DELQ_{i,t-j} + \alpha_{4,j} (\Delta DELQ \times ASVIN)_{i,t-j} + \alpha_{5,j} (\Delta DELQ \times RASVIS)_{i,t-j}\right) \\ + \sum_{m} \beta_m Controls_{i,t-m}^{m} + Year_t + \delta_i + \varepsilon_{i,t} \quad (7)$$

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 7 presents the regression results of Equation (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.

[Table 7]

5.5 Robustness check

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.[10] Specifically, if the values of each group indicate they can correctly answer the 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.

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

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:

$$
\begin{aligned}
Default_{i,t} = & \sum_{j=1,3} \left(\alpha_{1,j} ASVIN_{i,t-j} + \alpha_{2,j} RASVIS_{i,t-j}\right) + \alpha_3 HFL_i \\
& + \sum_{j=1,3} \left(\alpha_{4,j} (HFL \times ASVIN)_{i,t-j} + \alpha_{5,j} (HFL \times RASVIS)_{i,t-j}\right) \\
& + \sum_{m} \beta_m Controls_{i,t-m}^{m} + Year_t + \varepsilon_{i,t} \qquad (8)
\end{aligned}
$$

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 (9) are presented in Column *Full* of Table 8. The coefficients for ASVI, i.e., α_1 and $\alpha_2$in Equation (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), the results of which are presented in Column *Less Literacy* and Column *More Literacy* of Table 8. 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.

[Table 8]

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.

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:

$$NASVI_{i,t} = \frac{1}{3}\sum_{j=0}^{3} SVI_{i,t-i} - \frac{1}{12}\sum_{j=0}^{11} SVI_{i,t-j} \quad (9)$$

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 4.3 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). The results are presented in Table 9. 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 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.

[Table 9]

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 10 presents the new regression results based on Equation (4).

[Table 10]

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.

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 4.3 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) are presented in Table 11. 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 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 11]

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.

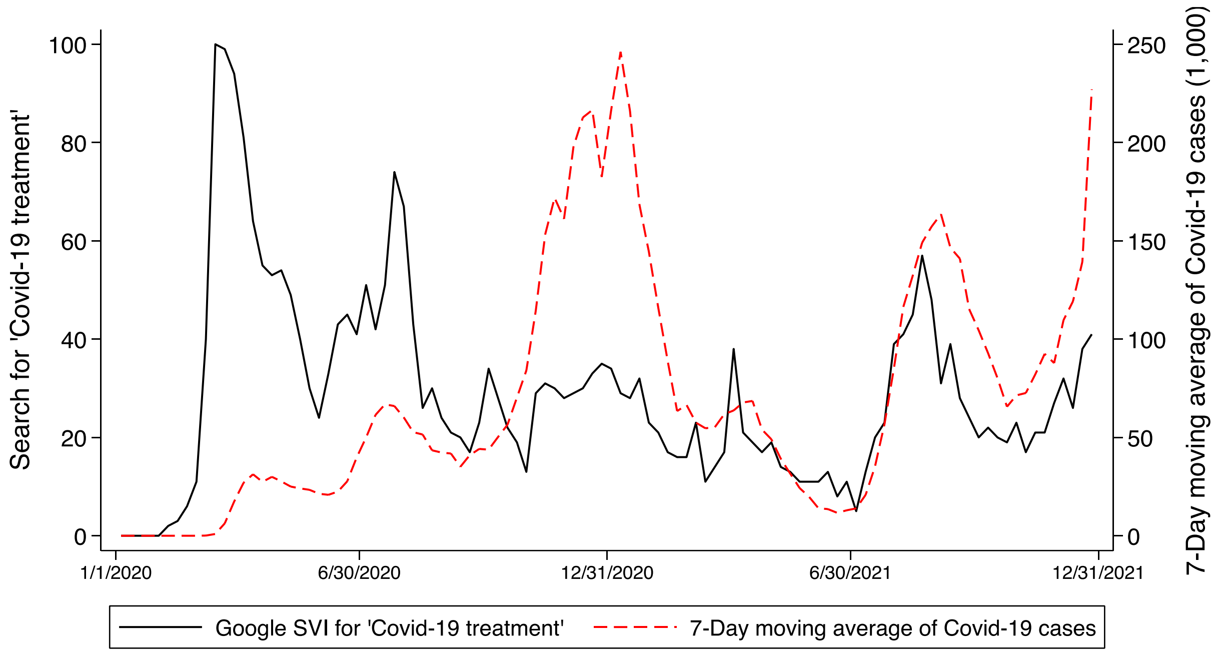


Figure 1: National search for "Covid-19 treatment" and the number of Covid-19 cases.

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

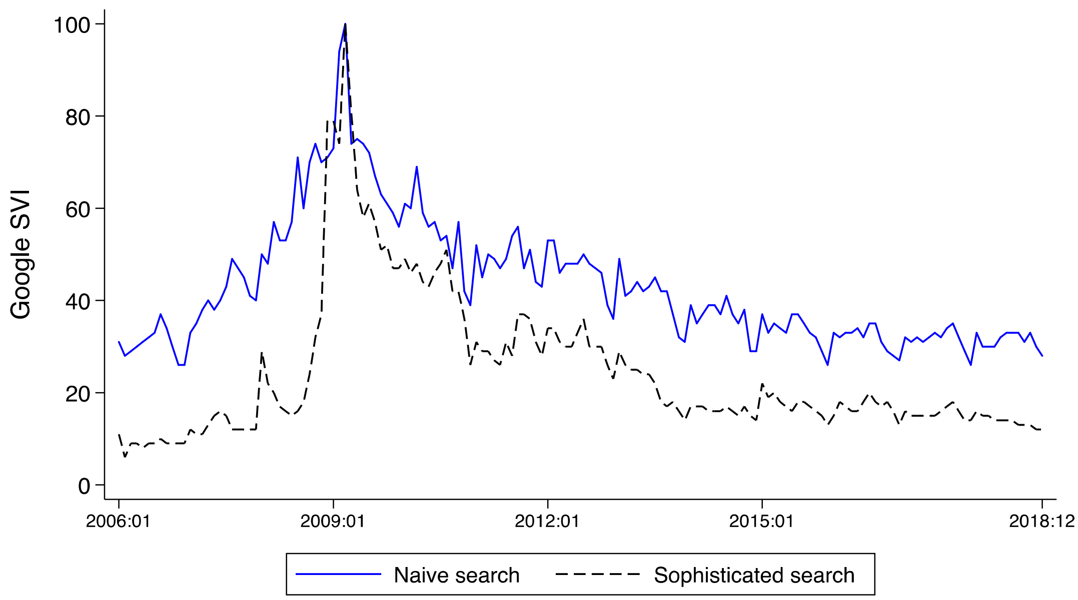


Figure 2: National naïve and sophisticated search activities.

Notes: This figure shows the trend of the original U.S. Google SVI data for the two joint search terms listed in Table 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.

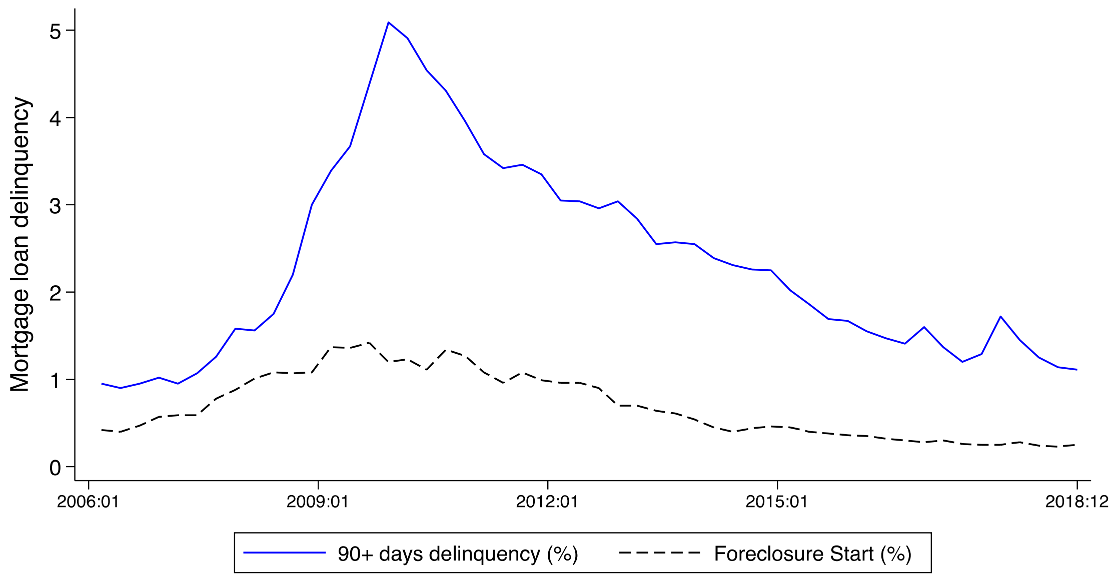


Figure 3: National mortgage loan delinquency.

Notes: This figure depicts the dynamics of two mortgage loan default indicators in the US. 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.

Table 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.

Table 2: Descriptive statistics.

Variables	Abbr.	N	Mean	Max	Min	Std. Dev.	Stationary test	Transformation	Geographic regions
Naïve search	ASVIN	2499	0.00	4.75	-4.13	1.00	-29.26***	original value	U.S. states
Sophisticated search	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
Unnemployment 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 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 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. *, **, 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.

Table 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.

Table 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)
$Dep.Var_{t-1}$	0.091***	0.087***	0.092***	0.708***	0.706***	0.703***
	(2.765)	(2.765)	(2.836)	(17.199)	(17.064)	(17.450)
ΔHP_{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)
$\Delta Income_{t-1}$	-2.727***	-3.001***	-2.724***	-1.082***	-0.994***	-1.084***
	(-3.046)	(-3.172)	(-3.063)	(-3.210)	(-2.978)	(-3.206)
$\Delta Unemp_{t-1}$	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)
$\Delta Loansum_{t-12 \text{ to } t}$	-0.249***	-0.247***	-0.247***	-0.085**	-0.082**	-0.091**
	(-5.069)	(-4.677)	(-5.025)	(-2.193)	(-2.167)	(-2.332)
$\Delta Subsum_{t-12 \text{ to } t}$	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.

Table 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.

Table 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)
ΔDELQ_{t-1}	0.149***	0.118***	0.154***
	(5.553)	(5.224)	(5.664)
ΔDELQ_{t-3}	0.068***	0.076***	0.065***
	(5.577)	(5.692)	(5.823)
$(\Delta\text{DELQ}\times\text{ASVIN})_{t-1}$	0.016	0.032*	
	(0.756)	(1.852)	
$(\Delta\text{DELQ}\times\text{ASVIN})_{t-3}$	0.020	-0.013	
	(1.640)	(-0.865)	
$(\Delta\text{DELQ}\times\text{RASVIS})_{t-1}$	-0.003		0.010
	(-0.173)		(0.831)
$(\Delta\text{DELQ}\times\text{RASVIS})_{t-3}$	-0.069***		-0.062***
	(-6.442)		(-6.260)
FS_{t-1}	0.759***	0.728***	0.754***
	(21.920)	(23.506)	(23.043)
ΔHP_{t-1}	-3.586***	-3.862***	-3.731***
	(-8.798)	(-9.907)	(-8.985)
$\Delta\text{Income}_{t-1}$	-0.777**	-0.740*	-0.831**
	(-2.102)	(-1.930)	(-2.164)
ΔUnemp_{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:

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

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.

Table 8: 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:

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

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.

Table 9: 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:

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

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.

Table 10: 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.

Table 11: Regressions on abnormal average SVI.

	ΔDELQ			FS		
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathrm{ASVINAVG}_{t-1}$	0.002	0.017**		0.018***	0.015***	
	(0.308)	(2.537)		(5.262)	(4.800)	
$\mathrm{ASVINAVG}_{t-3}$	-0.004	0.010		-0.001	-0.006**	
	(-0.525)	(1.486)		(-0.310)	(-2.194)	
$\mathrm{RASVISAVG}_{t-1}$	0.041***		0.041***	-0.009***		-0.001
	(5.915)		(5.931)	(-2.685)		(-0.344)
$\mathrm{RASVISAVG}_{t-3}$	0.026***		0.024***	-0.010***		-0.012***
	(4.252)		(4.905)	(-4.696)		(-5.357)
$\mathrm{Dep.\,Var}_{t-1}$	0.092***	0.095***	0.093***	0.716***	0.713***	0.709***
	(2.808)	(2.978)	(2.927)	(18.143)	(17.969)	(18.342)
$\Delta\mathrm{HP}_{t-1}$	-3.222***	-3.246***	-3.197***	-4.600***	-4.653***	-4.726***
	(-5.353)	(-5.358)	(-5.237)	(-11.301)	(-11.429)	(-11.157)
$\Delta\mathrm{Income}_{t-1}$	-3.421***	-3.529***	-3.408***	-1.251***	-1.211***	-1.320***
	(-4.356)	(-4.173)	(-4.433)	(-3.690)	(-3.588)	(-3.718)
$\Delta\mathrm{Unemp}_{t-1}$	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

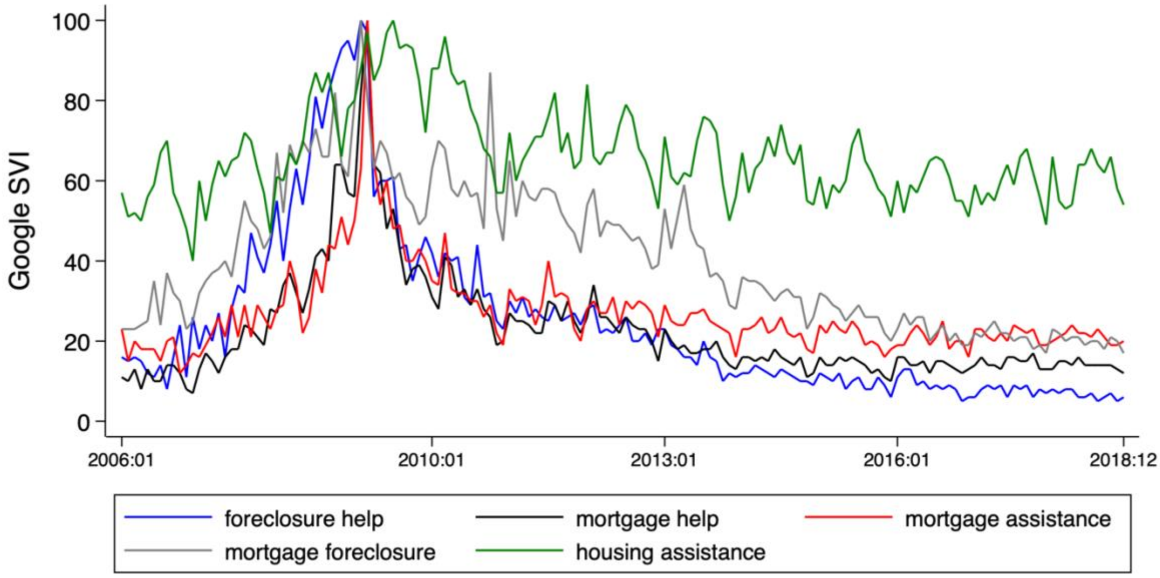
Notes: The table reports the results for regressions using abnormal SVI from averaged SVI data. We run the following regression using different mortgage default performance measures, either the change in 90+ days delinquency rate (ΔDELQ) or the foreclosure start rate (FS), as the dependent variables:

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

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.

Appendix

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

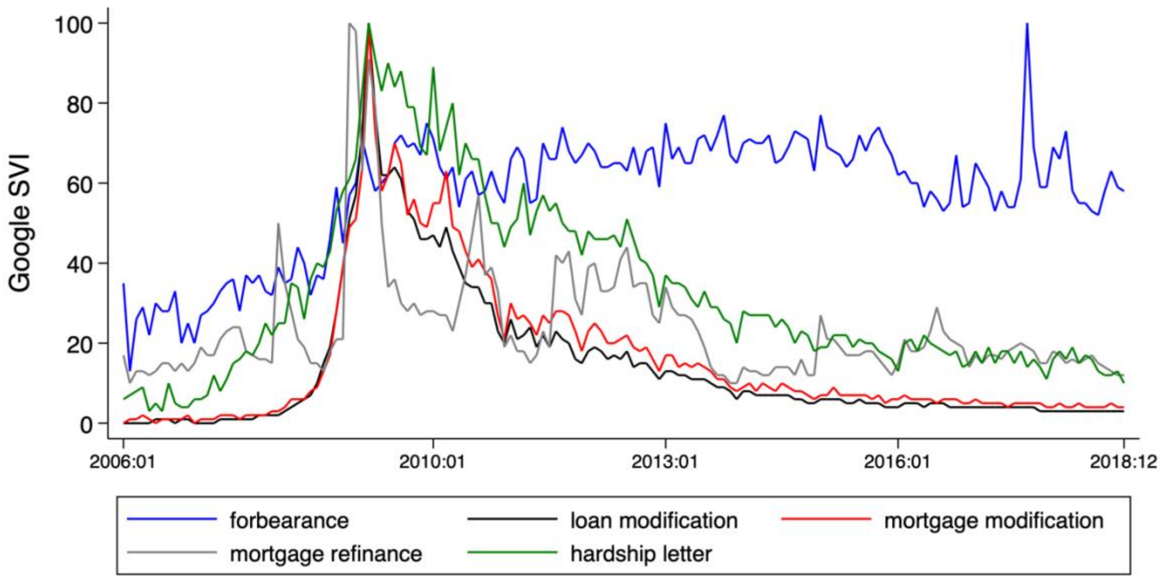


Figure A1: Dynamics of the Google SVI for independent search terms.

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

References

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.

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

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.

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)

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.

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

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

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

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

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.

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.

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

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.

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.

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.

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.

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.

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)

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.

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 Behavior*, 56, 103-118.

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 Behavior*, 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.

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 behavior, liquidity and prices in the housing market. *Real Estate Economics*, 46(2), 368-403.

Walhout, J., Oomen, P., Jarodzka, H., & Brand-Gruwel, S. (2017). Effects of task complexity on online search behavior 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.

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.