Detecting Mortgage Delinquencies with Google Trends[1]

Nikolaos Askitas*

Institute for the Study of Labor (IZA), Schaumburg-Lippe-Str. 5/9, D-53113, Bonn, Germany

Klaus F. Zimmermann

Institute for the Study of Labor (IZA), Schaumburg-Lippe-Str. 5/9, D-53113, Bonn, Germany and Bonn University

---

Abstract

Economic exuberance and hardship are strongly reflected by the housing market which is therefore the concern of much research. The timely monitoring of housing market conditions is often obstructed by insufficient and lagged data as is, unfortunately, often the case, especially in times of crisis. Our paper demonstrates that with elementary quantitative methods and properly drawn data it is possible to successfully keep tabs on housing market conditions in near real time. We evaluate the intensity of Google searches for "hardship letter". A hardship letter is nothing but a loan modification request a home owner writes to her bank. The demand for advice on how to write one should therefore be a leading indicator of the prevalence of mortgage delinquency. Searches for "hardship letter" are perhaps the earliest point in time a homeowner in financial distress externalises her deteriorating economic condition and Google Trends is the first instrument to detect the trend. We conclude that Big Data is the next big bet for Social Sciences.

JEL: C81, E65, G21, R31

Keywords: housing, hardship letter, loan delinquency, mortgage, National Delinquency Survey (NDS), financial crisis, recession, bubble, Google Trends

---

[1]An earlier version of this manuscript appeared as IZA Discussion Paper 5895.

*Corresponding Author

Email addresses: askitas@iza.org (Nikolaos Askitas), zimmermann@iza.org (Klaus F. Zimmermann)

1. Introduction

During its lifecycle (construction, sale, renovation, resale) a house causes demand for a wide range of raw materials and a diverse amount of labor. By the time it becomes a home it also generates consumer demand for a diverse array of home equipment. This is part of the reason that housing markets are important for national economies. The significance of a home and of homeownership for the lives of people is quite likely the prime force underlying this importance, but also the basis on which predatory lending, bad mortgaging, speculation, bubbles and other irrationalities occur. Nonetheless, conventional economic wisdom, not entirely unfounded, maintains that as the housing market goes so goes the rest of the economy for the US. This is easily seen to be the case in the recent great depression: the *"Great Surge in mortgage defaults"* (Dennis R. Capozza and Robert Van Order (2011)) was crowned by a financial and economic crisis which manifested themselves as such in the Fall of 2008.

The housing crisis was at the heart of the great depression. For an extended period, rising house prices had been the basis for a stable consumer demand. The drop of the value of houses and the resulting housing market collapse were detrimental for consumption and for the whole US economy.[2] Therefore, economic research needs to measure and understand its' nature at an early stage. Early data availability is one crucial challenge.

One way to get a sense for the size of the US housing market is to consider that the sample used by the *Mortgage Bankers Association* (`mortgagebankers.org`) to construct their National Delinquency Survey (NDS) includes about 44 millions loans held by "mortgage companies, commercial banks, thrifts, credit unions and others". A market of that size is highly important both in terms of its economic relevance and the amount of people it affects. According to the NDS, during the recent economic crisis, the total percent of delinquent loans went from 4.41% at the beginning of 2006 to 10.1% at the beginning of 2010 when within the NDS sample alone we had 4.5 million delinquent loans. In the same time interval quarterly foreclosures went from .98% to 4.63%.

In May of 2008, Angelo Mozilo, then Chairman of the board and CEO of a mortgage company called *Countrywide Financial*, made the news[3] by accidentally hitting "reply" instead of "forward" in response to an e-mail from a distressed homeowner in North Carolina. The homeowner, who

[2] However, as Molloy and Shan (2011) have shown, foreclosures on residential mortgages do not generate an economic burden strong enough to severely reduce housing consumption.

[3] http://articles.latimes.com/2008/may/21/business/fi-mozilo21

apparently was facing or anticipating an inability to make mortgage payments on time, had written a (hardship) letter to request a loan modification and emailed it directly to the Office of the President of Countrywide. Angelo Mozilo who apparently thought he was forwarding the email to one of his people commented it as follows:

> "This is unbelievable. Most of these letters now have the same wording. Obviously they are being counseled by some other person or by the internet. Disgusting."

Apparently, the homeowner who received the inadvertent reply with the comment made the email exchange public. *Countrywide Financial*, which in 2006 was the nation's *"largest mortgage lender"*,[4] was by then failing and was purchased by Bank of America[5]. Mozilo who at that moment must have been confronted with a prolonged rise in loan delinquencies saw it right. As more and more homeowners were entering financial hardship they sought help by writing loan modification requests to the banks holding their mortgages and the internet was their prime resource: a number of sites offered advice on the subject matter, *loansafe.org* being among the most prominent.

Our paper offers a novel way for the early detection of mortgage delinquency and hence of deteriorating economic conditions. More specifically we want to detect the likelihood of mortgage delinquency as early as the homeowners themselves sense upcoming difficulties. One way to do this would be to simply count Mozilo's letters. Since this is obviously impossible, we use *Google Trends* to find a proxy indicator that approximates their number. Such a letter is called a hardship letter and Mozilo had already seen that they are very much alike. This is of course because people search on the internet using terms like"hardship letter template" or "how to write a hardship letter". By counting the volumes of searches which contain `hardship letter` we can indeed detect loan delinquency in its early stages before any other indicator could possible pick it up. We also look at a number of other relevant searches such as `short sale`, `REO` ("Real Estate Owned") and `FHA` (Federal Housing Administration) to achieve a comprehensive big data approach to the monitoring of housing market.

The paper is organised as follows. In Section 2 we discuss the basic elements of an Internet Search Theory for the Social Sciences and present our data provisioning service: Google Trends. Section 3 contains descriptives of the housing market using a mixture of classical and search data

[4] `http://www.nytimes.com/2007/08/26/business/yourmoney/26country.html`
[5] See `http://www.marketwatch.com/story/bank-of-america-to-buy-countrywide-financial-for-4-billion`

while in Section 4 we show how a housing market practitioner is at all times better off using Google Trends for the monitoring of the housing market than a basic autoregressive model. We conclude with Section 5.

2. Internet Search: Concept and Practice

Online is a "place" where diverse socioeconomic activity takes place. From online banking, entertainment and shopping to informal education, social networking and information discovery the internet is a central resource for the lives of people in modern societies[6]. The study of internet activity is therefore of increasing importance for social sciences. Search is a vital component of this activity in that it indexes the enormous stock of online "documents" and makes them tractable for the user population. An individual's interest in certain documents (and not in others) is a function of the individual's state and so are search queries which are used to locate them. These queries are therefore utterances worth being investigated and their collection can be thought of as an involuntary, high-frequency and variable-frequency panel survey. The panel nature has tremendous potential, but is less likely to be usable in the near future on grounds of data privacy and protection[7].

Individuals performing internet searches are protected by anonymity and the privacy of their internet session. Such searches are therefore, involuntary and hence contain no interviewer effects.[8] This is a very important property in some contexts such as health studies (Askitas and Zimmermann (2011)). Protecting the privacy of individuals, in this context, is important and raises many interesting legal and ethical issues. These utterances come in a high but variable frequency with a panel nature generating advantages and challenges at the same time. Since the data inherits the internet's own geographic distribution it also has a huge cross-sectional potential.

Focusing on certain types of properly chosen queries, we may be able to successfully define proxies of the documents they locate and hence measure proxies of the individual's state. Reasonably

[6]Even in emerging economies where landlines and fast internet are not as prevalent wireless telephony is ubiquitous suggesting that the use of digital data for the monitoring of socioeconomic conditions would be a valid undertaking.

[7]As we know, in this post-Snowden era, players (such as the NSA) foreign to scientific investigation and hence without a directly legitimate societal mandate are also interested in internet activity and often end up having a more privileged access to such data.

[8]Internet activity data have been used before as a reliable data base for econometric investigations. See Askitas and Zimmermann (2009) for a seminal use of Google activity data for the analysis of unemployment data in the great depression. See also Kulkarni, Haynes, Stough and Paelinck (2009) for using Google activity data to predict house prices as well as McLaren and Shanbhogue (2011 Q2) for more recent efforts.

defined aggregate measures of intensity by which certain classes of searches are pursued may then be thought of as a sort of indicator of the degree of proliferation of a certain condition. In societies with high internet penetration, the reservoir of internet users is also a reasonably representative sample of the general population. The type of socioeconomic phenomena we may be able to detect range from emotional, psychological and physical health, epidemics, economic conditions, crime etc.

Finally, search represents only the demand side in this context. The supply of documents through internet platforms is equally important but it is not easy to get access to this type of data at the moment. Future research would benefit from a merger of supply side data with internet search data.

In this paper we use *Google Trends* data (see `http://www.google.com/trends/`). Since the summer of 2008, *Google Trends* allows a limited view into relative search volumes. For chosen regional, temporal, search term specific and Google category specific parameters, Google Trends will return weekly or monthly time series starting as far back as 2004. One can query *Google Trends* for searches which contain specific terms, disjunctions thereof as well as complements: "Jobs -Steve" will give you search intensity for the class of searches containing "Jobs" without those which also contain "Steve". Google divides all search in categories such as "Health" or "Automotive" and volumes of search classes can be obtained relative to a category (restricting "jobs" searches in the "Classifieds" category helps take out the noise coming from "Steve").

Comparative queries are possible: up to 5 search terms in a single geographic unit for any time interval or a single search term in up to five geographic regions in any time interval. Volumes are scaled and normalised as follows: All data points of the time series will be normalised by dividing the search term volume with the total search volume in the reference time interval (day, week, month). This means if the keyword K has K_i searches in the i-th reference time point and the total Google search volume therein is G_i then the measurement is $N(K_i) = K_i/G_i$. The latter will then be scaled by setting the maximum value equal to 100 and scaling the rest accordingly, i.e. the series based on Google Trends is then $I(K_i) = \frac{N(K_i)}{\max_i\{N(K_i)\}}100$. This means that two search terms queries in a comparative modus deliver relative volumes which are nonetheless comparable with each other. Finally, sampling is involved in generating these numbers and the results are robust for sufficiently high volumes. For low values results may be unreliable and caution should be applied.

An expected typical use of this type of data is to capture search terms which deviate from their mean and use semantics to relate it to some type of socioeconomic activity for which only low frequency, lagged measurements are available. This then allows us to get early, high frequency hints on breaking trends. It is unlikely that a certain keyword will have the same significance in the long run. For example if in ten years time one calls a hardship letter differently (because for example new legislation introduced some different terminology) then our current model will not be stable. However there will always be search terms that relate to the population's current state. An algorithm by which to operate therefore would be to monitor all search terms (viewing the search chatter as some type of unsolicited, involuntary, utterance) detect deviations from the mean and then apply semantics to look closer and find out what the searches are about. Once one gets the semantics right a target low frequency socioeconomic variable may be available which could be used to test the significance of the search deviation from its mean.

We now describe two example in order to enable the reader to better relate to this type of measurement. The first example illustrates the degree of disruption this type of data represents and has to do with the analogue library catalogue card of the past. Each library had some drawers in which every book in the library's inventory had a card with some basic metadata. If one could get frequent measurements on the titles of books checked out and their bibliographic metadata one would have been able to monitor what people read and perhaps forecast such things as how many engineers or mathematicians there will be 4-5 years into the future. This data were fragmented and hard to compile and analyse and that is the reason that we could not do things with them. With internet data and in particular Google Trends we have exactly this instrument at our disposal in a global scale. Since we only have relative measurements though a second example is due. Imagine being outside a soccer stadium and listening to the collective incomprehensible chatter of, say 50,000 spectators. We will never see a goal but we will be able to "hear" when one side senses danger or a scoring chance, when a missed scoring opportunity occurs, when a goal occurred or a violent foul or if the home team attacks or the visiting team. Any sports fan has a built in model which would allow them to tell what the score is at the end of a game of 90 minutes without having viewed the game.

3. Data Descriptives

In this section we will present our results by isolating some key internet searches that best capture the housing market developments in the time interval 2004-2013. This interval contains a time span "before" the current economic and financial crisis. The date we will use to split time into one segment before and one after the crisis is October 3 2008 on which the Troubled Assets Rescue Program (TARP) was signed into effect.

Figure 1 shows the variation of some key economic variables during this time. Clockwise from the top left we see the percentage of overdue loans, the total number of loans, the (housing) consumer confidence and housing prices all plotted agains the backdrop of Initial Jobless Claims. We notice that the labor market is of course highly seasonal and was slightly worsening during the years before t=TARP but most jobs were lost shortly after that. This demonstrates that it was not worsening labor market conditions which caused delinquencies to rise but most likely quite the reverse occurred: a bursting housing bubble which caught some "pundits and economists" with "their parameters down" (Sanders and Order (2011)) started an unstoppable surge of defaults which culminated with the crisis in the Fall of 2008.

What makes our approach valuable is that we can provide practically weekly monitoring of housing market conditions. The percentage of delinquent loans and foreclosures was increasing well before t=TARP. This poured houses into the market causing prices to drop and consumer confidence to rise. The mortgage industry continued to issue new loans up until t=TARP. This is the backdrop against which we now turn our attention to the reservoir of home owners keeping in mind that payment difficulties were increasing since as early as 2006.

"Some 69% of all Americans have used the internet to cope with the recession as they hunt for bargains, jobs, ways to upgrade their skills, better investment strategies, housing options, and government benefits. That amounts to 88% of internet users."[9]. It is hence to be expected that homeowners are actively using the internet.

The first thing a homeowner in financial dire straights might do to rescue their home is to request a modification of their mortgage payments. This is the case whenever there is hope that the hardship is only temporary. The lender has an interest in considering easing up a bit in order to avoid a foreclosure and secure the long term repayment of the mortgage. The letter a borrower

[9] http://www.pewinternet.org/Reports/2009/11-The-Internet-and-the-Recession.aspx

writes in this case is called a hardship letter. It is designed to explain the reasons for failing to make mortgage payments and why these reasons are of a temporary nature. Obviously, a hardship letter is written by a homeowner who has not given up on their home yet. If this is not the case and a foreclosure is imminent, a short sale is the next possible option which is milder than a foreclosure.

A short sale is *"a sale of a house in which the proceeds fall short of what the owner still owes on the mortgage. Short sales usually occur when the homeowner is facing foreclosure. Many lenders will agree to accept the proceeds of a short sale and forgive the rest of what is owed on the mortgage when the owner cannot make the mortgage payments. By accepting a short sale, the lender can avoid a lengthy and costly foreclosure, and the owner is able to pay off the loan for less than what is owed."*[10]

If push comes to shove the next stage of a home loan gone bad is a foreclosure, which is when a lender tries to recover some of the unpaid loan. Since a foreclosure is the step after a failed short sale the chances of it producing a sale are low and hence the property becomes an asset of the lender and is now what is known as *"Real Estate Owned"* property (REO). It is at this stage that a REO house begins to be attractive for new buyers. Such property is sold "as is" which may include ware and tear damages as well as damages from inhabitants in default.

Google search captures these stages fairly well. Searches containing `hardship letter` are made by homeowners who are getting ready to write one, while searches containing `short sale` are performed by those who want to find out what a short sale is; this may include the readers of our paper who want to find out more about short sales but it also includes the homeowners who are having trouble making due and need to consider their options. Searches for REO quite likely capture the demand for repossessed property and hence its supply. Therefore, we have enough "probable cause" to suspect that deviation of a reasonably defined volume of searches along these keywords from its known mean is probably a signal of the proliferation of adversity in the housing market.

Figure 2 shows that the weekly volumes of searches for `hardship letter` relate well to the percentage of delinquent loans and, in fact, may well be used to monitor market conditions. The searches are projected onto the Google category "Finance & Insurance / Credit & Lending" while the loan delinquencies are quarterly and taken from the National Delinquency Survey ((Mortgage

[10]Nolo's Plain English Law Dictionary: `http://www.nolo.com/dictionary/short-sale-term.html`

Bankers Association (2011))). Figure 2 quite impressively demonstrates the value of the internet data; not only do they predict reality well much earlier (with a lead of at least 6 weeks due to NDS publication lag), but also with more detail, since the internet data is weekly and hence provides a good nowcasting for loan delinquencies. This lead is important for identifying trends early, but the activity data also enables research to investigate the causes of change ahead of time.

At this point a note on the data is important. The series for `hardship letter` was constructed by drawing the weekly time series ten times from Google Trends (sampling was done at 13:56:36 on 30 Dec 2013 10) and taking a point-wise average of the ten draws. We then took 13-week moving averages which involve only lags so that forecasting value remains unaffected. The signal quality improves dramatically when we take point wise averages of multiple draws as can be seen by the prediction errors from single draws versus those from averaged ones. In the next section we will do some forecasting exercises with this time series.

Figure 3 shows the searches for `short sale` and captures the "epidemiological view of the housing market" from a different angle. As economic conditions become more uncertain and defaults, short sales and foreclosures become more common so do property repossessions and hence an increased interest in Real Estate Owned (REO) homes. This is exactly what Figure 4 shows. Both Figures 3 and 4, and the respective developments of internet activities searches for `short sale` and `REO` have indicated well the tensions in the housing markets and have "predicted" TARP.

The Federal Housing Administration (FHA) was created in 1934 (and it is part of the Department of Housing and Urban Development's (HUD) Office of Housing since 1965). When the FHA was created, *"the housing industry was flat on its back"*[11] as *"two million construction workers had lost their jobs"*, *"terms were difficult to meet for homebuyers seeking mortgages"* and *"America was primarily a nation of renters"* as *"only four in 10 households owned homes"*[11].

The FHA, *"provides mortgage insurance on loans made by FHA-approved lenders throughout the United States and its territories". "FHA mortgage insurance provides lenders with protection against losses as the result of homeowners defaulting on their mortgage loans"*. What makes FHA different is that *"unlike conventional loans that adhere to strict underwriting guidelines, FHA-insured loans require very little cash investment to close a loan. There is more flexibility in calculating household income and payment ratios. The cost of the mortgage insurance is passed along to*

[11] `http://www.hud.gov/offices/hsg/fhahistory.cfm`

the homeowner and typically is included in the monthly payment.[11] Figure 5 shows that FHA related Google searches were surging since early 2007. In our opinion, this surge indicates increasing uncertainty since it indicates that the industry is trying to draw in more and more weakly financed buyers. The figure shows that most of the spike is explained by FHA related searches also containing one or more of the keywords: `real, housing, connection, insurance, mortgage, requirements, loan, loans, rates, refinance, guidelines, approved, appraisal, appraisals, appraiser, homes,appraisers, condos, condo, appraisers` which allows us to see the intentions of the Google users.

4. Regression Results

In this section we perform an elementary model comparison exercise. We take the view of a practitioner such as a policy maker, (central or mortgage) bank analyst and we compare a basic "hardship letter" model against a basic autoregressive model at various times where they are simultaneously available[12]. We introduce the necessary symbolism in order to write down our models. Let y_t be the total mortgage delinquency at time t, where t is a quarter and let x_w be the relative search intensity for "hardship letter"[13], here w expresses time in weeks. We aggregate the weekly searches by taking 13-week (i.e. "quarterly") moving averages so that we now have "quarterly" searches with weakly lags. We thus can write $L_i.\overline{x}_t$ to indicate the i-th weekly lag of $\overline{x}_t$, the 13-week moving average of x_w at time t and $S_i.\overline{x}_t$ for the i-period difference. We can now write down two types of models:

$$\text{Model}A_j : y_t = ay_{t-j} + bS_4.y_{t-j} + c + \epsilon_t,\ j = 1, 2 \tag{1a}$$

$$\text{Model}HL_i : y_t = \overline{a}L_i.\overline{x}_t + \overline{b}S_{52}.L_i.\overline{x}_t + \overline{c} + \overline{\epsilon}_t,\ i = 9, 6, 3, 0 \tag{1b}$$

Since the MBA National Delinquency Survey is published 6-7 weeks after the completion of each quarter in forecasting y_t we can compare model (1b) against model (1a) when $i = 9$, $j = 2$

[12] Our aim is not to exhaustively investigate all currently available forecasting technology against our problem but to demonstrate that data chosen in the way we discuss outperforms the next best option in a basic way. This (and the lack of enough data points) is the reason we do not look into such things as MIDAS regressions and the like.

[13] We took the search intensity for the search term "hardship letter" in the US restricted to the category "Finance/Credit &Lending". We drew the data at 13:56:36 on 30 Dec 2013 10 successive times and formed a new time series by taking point wise averages. Google Trends data is sampled from the total so there may be significant variation at different draws. The differences are inversely proportional to the search volumes. Forming the point wise average strongly improves the results.

or $i = 6, 3, 0$, $j = 1$. We chose these values because for $i = 9$ a practitioner would have model (1a) only for $j = 2$ and her disposal. Model (1b) is available on a weekly basis and we chose $i = 9$, the mid point between when y_{t-2} is known and the publication of y_{t-1}. At $i = 6$ we have the publication of y_{t-1} and hence we can compare models (1a) and (1b) as soon as they become available ($j = 1, i = 6$) in the midpoint between the publication of y_{t-1} and the completion of the quarter ($j = 1, i = 3$) and upon the completion of the quarter ($j = 1, i = 0$). We compare the models in terms of mean squared percentage error (MSPE), average adjusted R^2 and average AIC by running rolling one period forecasts starting at quarter $t = 10$ i.e. the second quarter of 2008. We summarise the comparison in the table below from which we clearly see that a practitioner would be better off using google data than autoregression.

	A_2	HL_9	A_1	HL_6	HL_3	HL_0
MSPE	1.579	.772	.904	.526	.387	.367
avg AR^2	.856	.964	.932	.974	.412	.982
avg AIC	24.545	13.301	16.349	5.653	.978	-3.483

5. Conclusions

The internet is a place where much socioeconomic activity takes place and that makes it an important, and largely untapped, resource of information for scientific analysis. Sorting, searching, accessing, and discussing internet documents reveals important information about the lives of people and their activity. The internet as a data source has some very interesting properties: it has a high frequency, contains no interviewer effects and it is available in real time.

This paper demonstrates that weekly monitoring of housing market adversity in the US is feasible by comparing hardship letter searches with delinquencies of housing loans. In fact the hardship letter searches using Google Insights predict delinquencies quite well. Furthermore, we also show that searches for the services of the Federal Housing Administration (FHA), for info on short sales, and Real Estate Owned property, also captured various aspects of the housing market in the crisis very well.

Therefore, the internet search data is not only a valuable basis for revealing early new trends, it can also be an invaluable source for behavioural research. The housing market is a well suited example where the study of internet data can be successful, since home owners are likely to use the internet intensively. Housing professionals of all types, mortgage banks but also policy makers who need timely information on economic trends will benefit from this new data source.

References

Askitas, N., 2011. Weeklyclaims: Stata module to get weekly initial jobless claims from the us dept. of labor. Statistical Software Components, Boston College Department of Economics.

Askitas, N., Zimmermann, K.F., 2009. Google econometrics and unemployment forecasting. Applied Economics Quarterly 55, 107–120.

Askitas, N., Zimmermann, K.F., 2011. Health and Well-being in the Crisis. IZA Discussion Paper Series No 5601.

Dennis R. Capozza and Robert Van Order, 2011. The great surge in mortgage defaults 2006-2009: The comparative roles of economic conditions, underwriting and moral hazard. Journal of Housing Economics 20, 141 – 151. Special Issue: Housing and the Credit Crunch.

Google, 2008. Google Insights. `http://www.google.com/insights/search/`.

Kulkarni, R., Haynes, K.E., Stough, R.R., Paelinck, J.H., 2009. Forecasting Housing Prices with Google Econometrics. SSRN eLibrary.

McLaren, N., Shanbhogue, R., 2011 Q2. Using internet search data as economic indicators. Quarterly Bulletin (`http://www.bankofengland.co.uk/publications/quarterlybulletin/qb110206.pdf`).

Molloy, R., Shan, H., 2011. The Post-Foreclosure Experience of U.S. Households. Federal Reserve Board Finance and Economics Discussion Series .

Mortgage Bankers Association, 2011. National Delinquency Survey. `http://www.mbaa.org/ResearchandForecasts/ProductsandSurveys/NationalDelinquencySurvey.htm`.

Sanders, A.B., Order, R.V., 2011. Introduction to the special issue on housing and the credit crunch. Journal of Housing Economics 20, 66 – 67. Special Issue: Housing and the Credit Crunch.

Standard & Poors, 2009. S &P/Case-Shiller Home Price Indices Methodology. `http://www.standardandpoors.com/`.

University of Michigan and Thomson Reuters, 2011. Surveys of Consumers. `http://www.sca.isr.umich.edu/`.

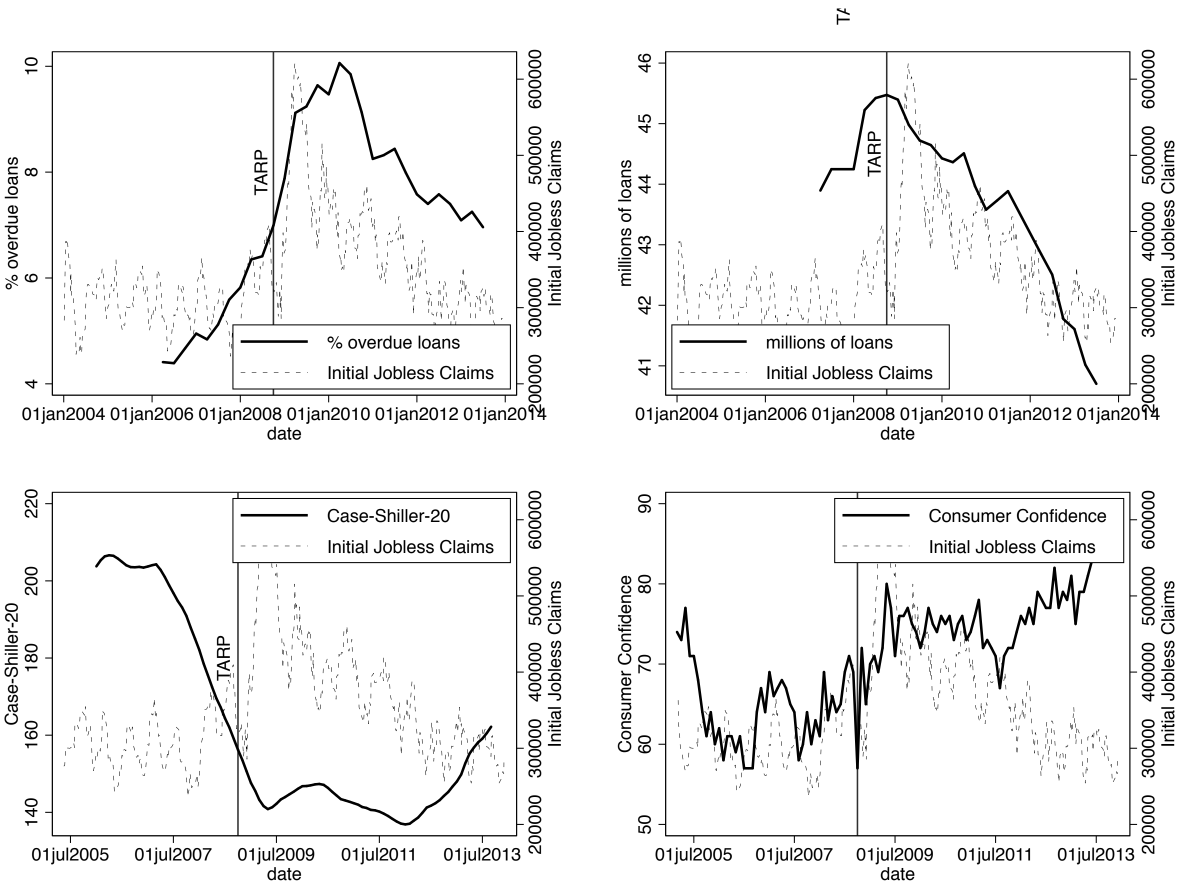

Figure 1: The US housing market 2004-2013.
Sources: Initial Jobless Claims are collected from the US Dept. of Labor. The percentage of loans overdue and the total volume of loans are from Mortgage Bankers Association (2011). House prices are expressed by mean of S&P/CS-20 from Standard & Poors (2009) and consumer confidence is obtained from University of Michigan and Thomson Reuters (2011).

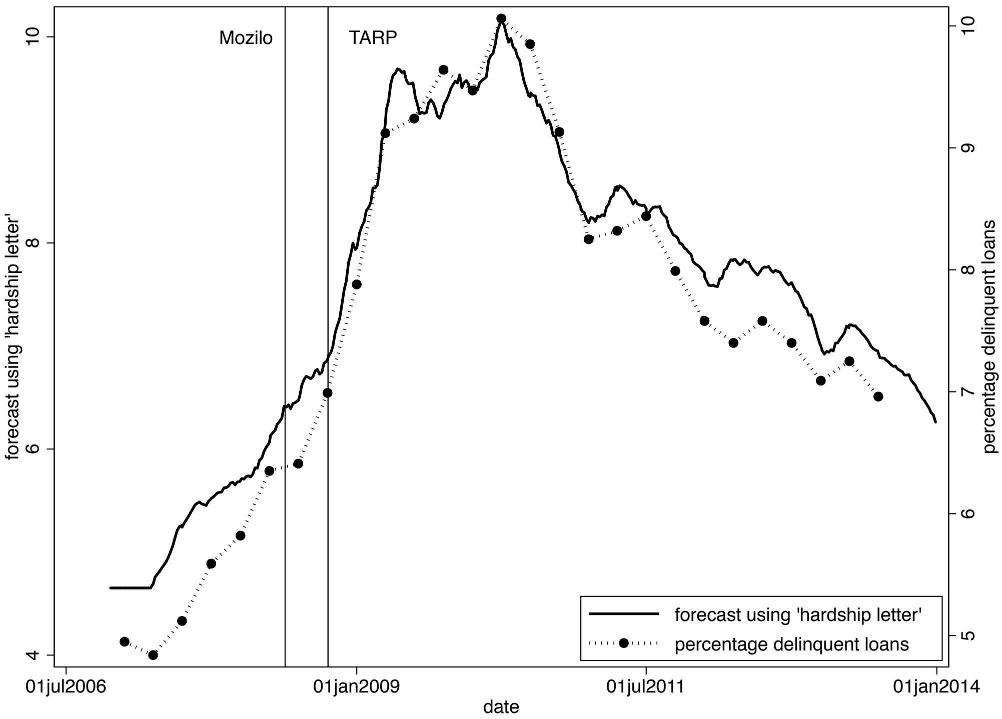

Figure 2: Searches for hardship letter predict mortgage delinquencies.
Sources: Searches for "hardship letter" (Google (2008)) are projected on the Google Category "Finance / Credit & Lending" and are compared to loan delinquencies (Mortgage Bankers Association (2011)). We take a 13 week moving average smoothing of the hardship letter searches without forward terms.

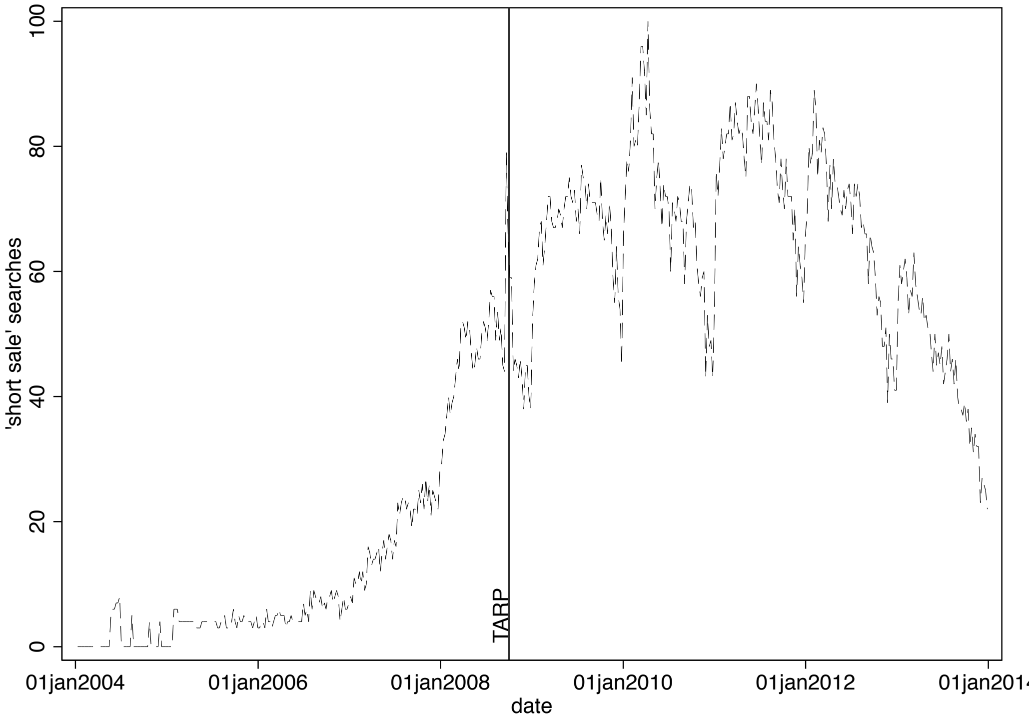

Figure 3: Searches for 'short sale'

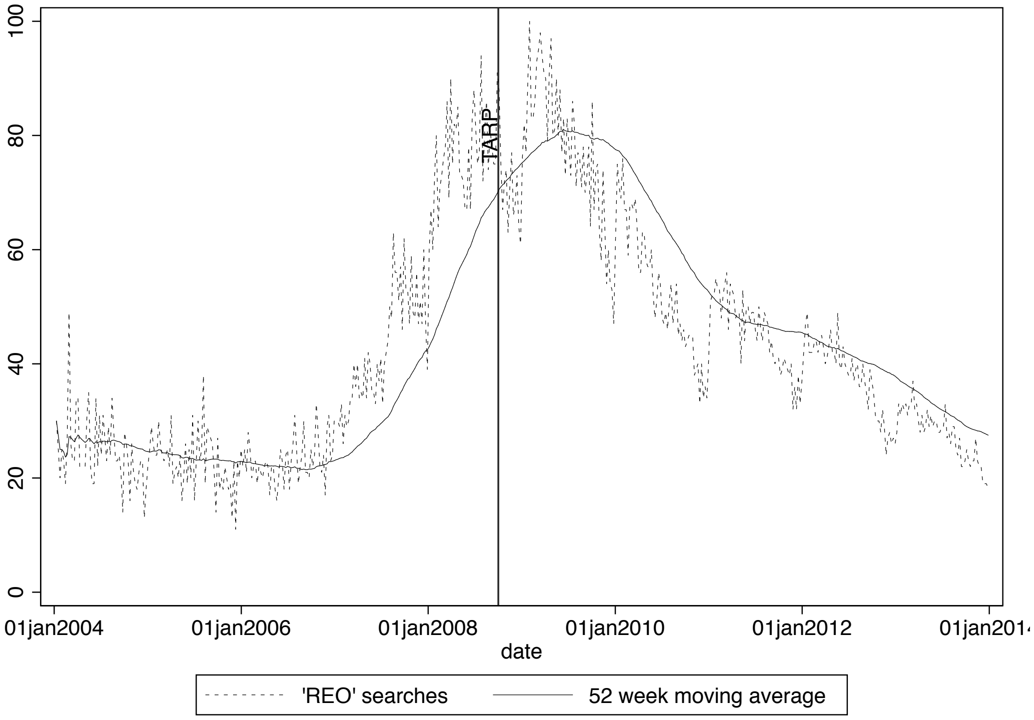

Figure 4: Searches for 'REO' in the Google Category "Real Estate"

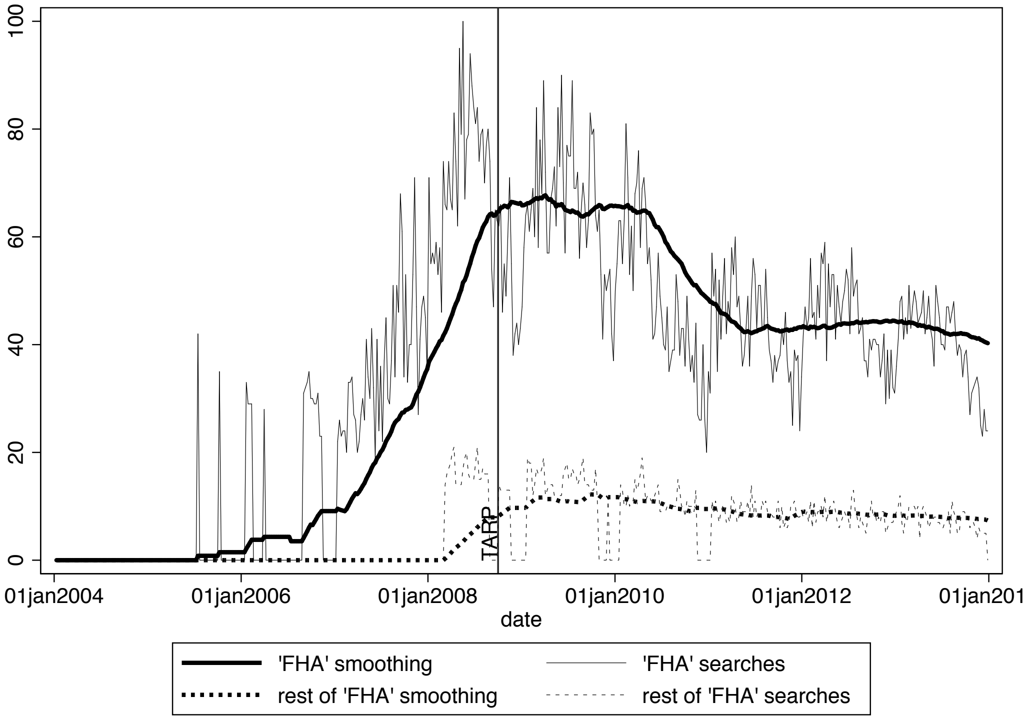

Figure 5: Weekly searches for "FHA" and smoothings.
By "rest of 'FHA' searches" we mean FHA searches which do not contain any of the words: real, housing, connection, insurance, mortgage, requirements, loan, loans, rates, refinance, guidelines, approved, appraisal, appraisals, appraiser, homes,appraisers, condos, condo, appraisers. This is a good example where we know what people aim at when they search for FHA.