Gauging Public Interest on Mortgage Prepayments: Analyzing Google Searches and Tweets Against Conditional Prepayment Rates

A summary

Submitted to the Central European University in conformity with the requirements for the degree:

Master of Finance

Capstone Project – Ansar Kokbayev

2000385

CENTRAL EUROPEAN UNIVERSITY

Supervisor: Dr. Zoltan Papp

Contents

Chapter 1: Introduction

1.1 Research Framework

This research is conducted with the goal to improve the existing Mortgage Prepayment models using internet activity related to the topic expressed by users in basic web searches and posts on twitter. Establishing a reasonable relationship between internet activity and public discussions of interest rate changes and mortgage refinancing might be a useful indicator of fluctuating mortgage prepayments.

1.2 Used tools and instruments

Web-scraping of the tweets was done using the open sourced *tweepy* library that uses the MIT license and was made as a research tool. This library has an ability to retrieve all the related data that comes to the tweets such as retweets, likes and replies (comments). All the related data manipulation was conducted using Python and the quantitative analysis in R. Google search queries and their popularity were retrieved from the official trends.google.com source.

Chapter 2: Public Interest

2.1 Introduction

Using the social media web scraping and search engine analytics, this research has measured public interest in such topics related to mortgage refinancing as: interest rates, mortgage loans, mortgage refinancing and others. Such factors as mentions, “likes”, searches and overall trends were used to gauge the levels of public interest. This research inherently relies on an assumption that individuals spend at least some time before making such financially significant decisions such as mortgage refinancing. Having said that, the moment

when an individual spends time on the internet researching and entering discussions on social media about the prospect of refinancing is the main incentive to conduct such research in the first place. From this moment on, "public interest" term will be used exclusively to describe this process of researching and discussing mortgage refinance related topics online.

2.2 Measuring Public Interest

During the Twitter web-scraping, over 8 million tweets that contain the following words related to mortgage refinancing were scraped from January 2007 to May 2021: *home refinancing, interest rate, mortgage loan, mortgage rate, mortgage refinance, mortgage refinancing, refinance.* All of the related characteristics of the tweets such as replies, retweets and likes were scraped as well.

As for the google web-search part, Google Trends analytics platform was used that analyzes the popularity of these search queries: *mortgage, mortgage refinance, mortgage rate, mortgage loan, refinancing, interest rate.* Figure 3 shows the web-search trends of the mentioned queries and their juxtaposition with the inverse 30-year mortgage rate in red.

Chapter 3: Empirical Results

3.1 Relationship between Mortgage Prepayments

Having established that the monthly mortgage prepayments being manifested as the BCPR, the immediate relation between the BCPR and the public interest must have been demonstrated. The correlation analysis has revealed that there is a 0.82 correlation coefficient between the web-search of a keyword "refinancing" and the actual BCPR.

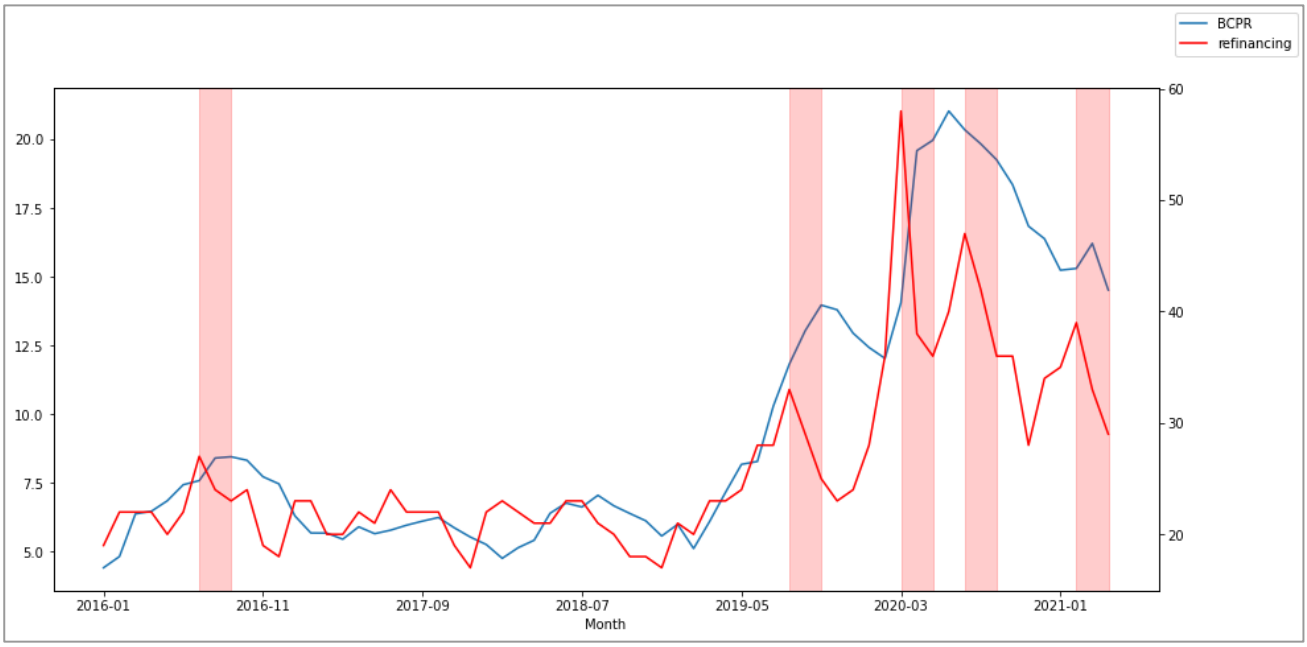

Figure 1 – Web-search of the keyword: "refinancing" and a highlighted two-week lag period between BCPR

If we look more closer to the timing of those google web-searches on Figure 1 above, an average of at least two weeks is detected between the web-search highs and BCPR highs. This observation leads to conclude two things: first is that it takes approximately two weeks on average to observe the effects of unusually high web-searches on the actual CPR. Second is that public interest in the topic is not a result of actual mortgage prepayments trends, but vice-versa, making predictions more realistic.

3.2 Regression Analysis

The purpose of the regression model in this research was to find reliable and significant evidence that the identified variables can be used to predict the movement of BCPR.

Chapter 4: Conclusions and Recommendations

Out of all the mortgage refinancing related web-searches and tweets, only a few were identified as statistically important with the help of regression models. Users interacting with the internet through google searches and twitter posts/discussions left clear patterns that can be easily used for more precise and fast mortgage prepayment modelling. This research has identified that the popularity of “refinance” search-query spikes on average two-weeks ahead of increases in mortgage prepayments measured using the BCPR by Fannie Mae. Technically, as simple as monitoring weekly google search trends can be used to anticipate extreme volatility in BCPR. Tweets mentioning mortgage rates are among the most popular, and most importantly, the most reliable in gauging public interest in the topic.

Finally, the regression model has demonstrated that the variables covered can be reliably used as leading indicators in mortgage prepayment models along with other classic economic variables such as inflation rates, mortgage rates and GDP.