Applied Macroeconomics: Micro Data for Macro Models

Homework 2

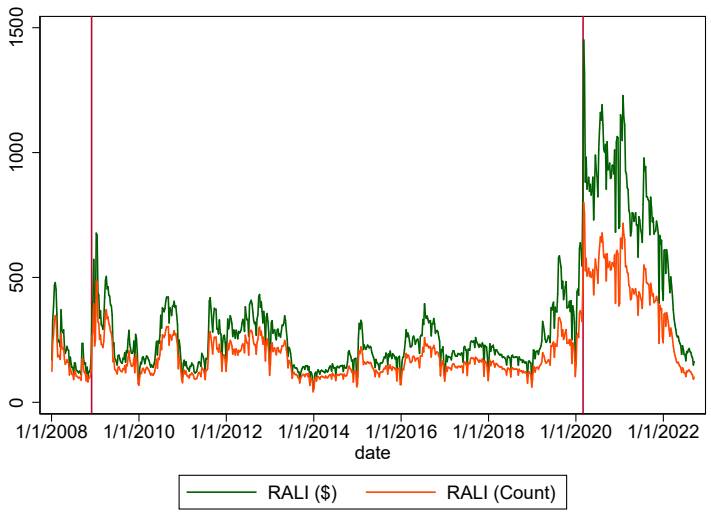
Regional Heterogeneity and Refinancing during COVID

Due: October 27th by 4:00 p.m.

In this homework, we will explore the recent refinance boom of 2020 and 2021 (Figure 1). During this period, low interest rates combined with robust house price growth contributed to a strong mortgage origination market. We will use public micro-data to understand how the refinance channel varied across regions and contributed to stimulating the economy post COVID.

Further, we will consider how the COVID period compares to the Great Recession by replicating some of the analysis in Beraja et al. (2018). In particular, you will compute refinance propensities by metropolitan area and study how these vary with house price growth and housing equity distribution.

Figure 1: Fannie Mae's Refinance Application-Level Index



Technical goals of this assignment:

- For you to get experience working with large datasets effectively. For this reason, we will require you to work on the university server (Mercury for Booth students and Athens/Acropolis for Econ).
- We also want you to figure out how to search for public data and codebooks, extract relevant variables, and how to merge them together.

Other:

1. We expect people to work on this project independently (so everyone can learn the required skills).
2. Olivia will be providing more guidance in her session on October 14th.
3. *Software*: You can use any software (Stata, R, etc). You will need to submit your code and a write-up with your plots, tables, and discussion.

1 Instructions

Download HDMA data to the server

Download the HDMA National Loan Level Dataset for 2020 and 2021. These are large datasets (∼20GB each) so you will need to download and work with these in the *scratch space* of the server. This is the recommended place to store temporary job files or intermediate output. To get your started, we will provide you with a shell script that will:

1. Create a working directory for this assignment in your scratch space.
2. Download the HDMA dataset directly from the public website and unzip it.

Note: You will need to open the shell script, understand it, and edit it such that it will run in the server you are using. We wrote it for Midway, but you will need to run it in Mercury or Athens. Olivia will be available to help you throughout, but we want you to figure out how to use the server and submit code.

Clean the HDMA data

You will want to clean and filter the data so it is more manageable. Find the HDMA codebook online and compute the following:

1. Filter the loan records to newly originated loans (Hint: action taken)
2. By MSA: Compute the number the number of refinancing loans (Hint: loan purpose), and the number of cash-out refinancing loans.[1]
3. By MSA: Compute the mean and median of the loan-to-value ratio. This is a measure of the leverage level (and therefore of housing equity).[2]

Compute refinancing propensity

From the 2019 ACS 1-year estimates (https://data.census.gov) download the number of occupied housing units by MSA and the number of owner-occupied housing units with a mortgage by MSA.

Compute the refinancing and cash-out propensity by MSA and by year (2020 and 2021) by dividing the number of new refinancing/cash-out loans by the stock of mortgages.

[1] A cash-out refinancing classifies as a refinancing loan; i.e. the total number of refinancing loans should include the cash-out loans.

[2] This is a measure of leverage at origination of new loans, ideally we would look at equity for the stock of loans.

Zillow price data

Download the month Zillow price data. Compute a 5-year price growth, 2015 to 2020 for 2020 and 2016 to 2021 for 2021. *Note*: You will need to do download the Zillow crosswalk to match the region IDs to CBSA IDs and be able to merge the price growth data to the refinancing propensities.

2 Analysis

House price growth

Produce a scatter plot of regional refinancing propensities vs. 2-year house price growth. Both for general refinancing and for cash-out refinancing. Additionally, run an OLS regression of refinancing propensities at the regional level on house price growth weighted by number of housing units.

How do you interpret these results? How does this compare to the great recession results documented in Beraja et al. (2018)?

Housing equity

How does housing equity varies with house price growth?

Run an OLS regression of refinancing propensities at the regional level on median leverage weighted by number of housing units.

How do you interpret these results? How does this compare to the great recession results documented in Beraja et al. (2018)?

Data and others

What potential issues do you see with using this type of public data (HDMA) instead of the more detailed CRISM data used in the paper?

What other forces could explain the regional heterogeneity we observe in refinance propensities during COVID?

References

Beraja, Martin, Andreas Fuster, Erik Hurst, and Joseph Vavra (2018). "Regional Heterogeneity and the Refinancing Channel of Monetary Policy*". In: *The Quarterly Journal of Economics* 134.1, 109–183. eprint: https://academic.oup.com/qje/article-pdf/134/1/109/27494598/qjy021.pdf. https://doi.org/10.1093/qje/qjy021.