**Estimating Bunching**

Shenghan Zhao

ECON 613

Department of Economics, Duke University

Dr. David Berger

Mar 13, 2021

**Abstract**

This project provides an estimate of bunching mass across different groups with different characteristics in response to conforming loan limit—the maximum loan size eligible for purchase by Fannie Mae and Freddie Mac. It also conducts an estimate of individual’s probability to fall into the strictly dominated region.

**1 Introduction**

In the U.S., homes and mortgage debts are two single largest items on most households’ assets and liabilities. The Federal Housing Finance Agency (FHFA) sets a conforming loan limit (CLL). If a household is purchasing a house valued near the CLL, there are very strong and salient incentives to only borrow just below it. This creates a notch in one’s budget constraint when taking a mortgage, and a bunching effect that people generally avoid taking out loans above the CLL and instead bunch to the limit at least theoretically (DeFusco and Paciorek, 2017). Under this setting, it is interesting to estimate the bunching mass, and find whether there are correlations between bunching estimates across groups and their sex, age, ethnicity. We will regress our bunching estimates on these characteristics later in our project. We will start with a homogenous, frictionless model and address these heterogeneity issues along with frictions in our project later.

**1.1 Research Goal**

We will first estimate the bunching mass, and run a regression of our bunching estimates on sex, age, ethnicity. And then discover about frictions and heterogeneity that prevents people from bunching.

**2 Literature Review and The Notch Model**

Defusco and Paciorek posit that households live for two periods, assuming that each household purchase one unit of house in the first period at an exogenous price . Households can finance their purchase with a mortgage , which should not exceed the value of the house . The interest rate is given by and does not depend on the mortgage amount in the baseline model. In the second period, Housing is liquidated, mortgage is paid off, and households consume all of their remaining wealth. The households’ problem now become to be the maximization of their lifetime utility by choosing non-housing consumption in each period, denoted by and . Defusco and Paciorek also define the income in the first period to be , and a discount factor . In general, the household solves the following problem:

|  |  |  |
| --- | --- | --- |
|  |  | (1)  (2)  (3)  (4) |

With the budget constaint:

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

Defusco and Paciorek linked the earnings response to the amount of bunching with the following equation:

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

where the approximation assumes the counterfactual density is constant on the bunching segment . Since the bunching mass, or the total bunching B, can be estimated simply by plotting the data we can then estimate the bunching interval, , by equation (6) after estimating the counterfactual distribution .

**3 Data and Research Design**

**3.1 Estimating the Counterfactual Distribution**

To estimate the counterfactual distribution, first take the logarithms of the loan amounts and center them at the CLL so that differences are shown in percentage terms. They then group those normalized loan amounts into bins centered at the values , where , and count the number of loans in each bin, . Note here is the excluded region around the conforming loan limit. After defining the excluded region, the following regression is fitted:

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

And the estimated counterfactual number of loans in each bin is calculated along with bunching and missing mass as:

|  |  |  |
| --- | --- | --- |
|  |  | (8) |
|  |  | (9) |

As we introduced before, in order to have an estimation of the bunching interval, , we need to have an estimation on bunching and the counterfactual distribution. Bunching is estimated following equation 9, and the counterfactual distribution is estimated by .

**3.2 Data**

We will be using CoreLogic data to estimate the counterfactual distribution and the bunching interval. HMDA data will be used afterwards to correlate bunching mass estimates with different sex, age, and ethnic groups. A simple OLS regression is our current plan. We will explore if OLS reveals some potential problems.

Since the real-world settings are naturally allowing for heterogeneity and frictions that prevent people from bunching. We will then take a look at the strictly dominated region, which should be empty under homogenous and frictionless assumption, and check the characteristics of individuals locate there, and use a probit model to predict individual’s probability of locating there.

**4 Plan of Research**

|  |  |
| --- | --- |
| Estimation of Counterfactual Distribution | April 8th |
| Estimation of Bunching Interval and Bunching Mass | April 9th |
| OLS Regression of Bunching on Groups’ Characteristics | April 11th |
| Look at Strictly dominated Region and Brainstorming | April 15th |
| Probit Model to Predict Probability of Locating at Strictly Dominated Region | April 20th |
| Write-up | April 26th |

**Reference**

DeFusco, A.A., & Paciorek, A. (2017). The Interest Rate Elasticity of Mortgage Demand: Evidence from Bunching at the Conforming Loan Limit. *American Economic Journal: Economic Policy,* 9(1): 210-240. <https://doi.org/10.1257/pol.20140108>

Kleven, H. (2016). Bunching. *Annual Review of Economics.* 8:435-464.

https://doi.org /10.1146/annurev-economics-080315-015234.