

This document describes the replication materials for the paper “Estimating Adjusting Frictions Using Non-Linear Budget Sets: Method and Evidence from the Earnings Test.” This paper uses restricted access data from the Social Security Administration (SSA) and the Longitudinal Employer-Household Database (LEHD).

The replication code consists of five parts:

1. Code that we ran at SSA, deriving some results from the microdata and also generating aggregate data (counts at the age-year-rounded earnings bin level).
2. Stata code that we ran on the aggregate data, producing the bunching estimates and earnings distribution figures.
3. Matlab code that we ran on the aggregate data and bunching estimates, producing the structural estimates reported in the paper.
4. Matlab code to do the policy simulations
5. Code that we ran in Census Research Data Center, generating results from the LEHD.

In addition to the replication code, we provide the aggregate data. The replication materials therefore permit all interested researchers to replicate all our results derived from the aggregate data. They also permit researchers with access to the SSA or LEHD data to replicate our results from those datasets.

### **1. Re-running the code to generate the raw data and results from SSA**

The programs to run at SSA are located in codeForSSA.

To generate the aggregate data underlying the bunching figures, researchers should modify the opening lines of mainRR2 to point to the relevant directory. The program will call subprograms to load the raw MEF data, format it, and output aggregate data that underlies figures 2 and 4.

To generate Figures 5 and B1, and Table 1, researchers should modify the opening lines of mainSSA to point to the relevant directory. The program will call subprograms to load the raw MEF data, format it, and generate these figures and tables. It also outputs the aggregate data that underlies Figures 7 and B2, the structural estimation, and Table B5.

### **2. Replicating the bunching figures and tables**

Figures 2a and 2b, 4a and 4b, and Tables B1, B2, and B5 can be replicated from the aggregated data. The programs to replicate the figures and tables are located in bunchingCode. To replicate the bunching figures, researchers should take the following steps:

1. Modify the global directory defined on line 20 of mainBunch.do so that it points to the location of the replicationPackage. Modify the location of figdir and tabdir, if desired.
2. Run “mainBunch.do”

The figures will be saved in replicationPackage/figures. The tables will be saved in replicationPackage/tables.

### **3. Replicating the elasticity and adjustment cost estimates**

To replicate tables 2-5, B3, and B4, run matlab/replicateStructuralEstimates. The results will be saved in a diary, output/tables/structuralEstimatesDATE.log, where DATE is the date the file was run.

#### **4. Replicating the policy simulation, Table B6.**

Run matlab/taxReformSimulation. The results will be saved in a diary, output/tables/tableB6.

##### **Overview of data**

We provide six datasets of aggregated data. Each data set consists of probabilities or counts of observations at the level of earnings bin by age. The data sets are:

1. dens800: Pools the main sample, using data for 1990-1999, in \$800 bins of distance to exempt amount.
2. Dens400: Pools the main sample, using data for 1990-1999, in \$400 bins of distance to exempt amount.
3. Dens1600: Pools the main sample, using data for 1990-1999, in \$1600 bins of distance to exempt amount.
4. dens800\_8399: Pools the main sample, using data for 1983-1999, in \$800 bins of distance to exempt amount.
5. s1Claim\_9099: Pools the sample of claimants, using data for 1990-1999, in \$800 bins of distance to exempt amount.
6. S1Born13\_8399: Pools the main sample but limited to people born in January-March, using data for 1983-1999, in \$800 bins of distance to exempt amount.

We also provide two time series of estimated bunching amounts:

1. yearStatsDensPoly\_7802\_6668\_800\_s1Main: Estimating bunching amount in each year 1978-2002, pooling 66-68 year olds, based on our main sample and specification.
2. yearStatsDensPoly\_5509\_6264\_800\_s1Main: Estimating bunching amount in each year 1955-2006, pooling 62-64 year olds, based on our main sample and specification.

#### **5. Re-running the code to generate results from the LEHD**

The program to run at the Census is located in codeForCensus.

To generate Figure 3, researchers should modify the opening lines of etearningschangebyage to point to the relevant directory. The program uses a sample generated by randomly keeping the full LEHD earnings history of 20 percent of individual identifiers in the full LEHD data. Running the code will generate Figure 3.