Econ 230B
Spring 2021
Emmanuel Saez
Gabriel Zucman
GSI Felipe Lobel, lobel@berkeley.edu

Problem Set 2 Solution

1. Mobility of High Income US Taxpayers across States

The goal of this exercise is to estimate the mobility of high income US taxpayers across US states due to variation in state income top tax rates across states and over time. High income US taxpayers are defined as tax filers reporting Adjusted Gross Income (AGI) above $1m.

a) Find online information on the state top income tax rates across all states for **2017** incomes. List the five states with the highest top tax rates (group T) and the five states with the lowest top rates (group C) along with the top tax rates in those 10 states. (NOTE: do not exclude zero tax states, if you have ties, keep the largest states in terms of population to have exactly ten states in each group).

The first column of Table 1 does this:

Table 1: Tax rates

	2017 Rankings		2017-2001 Rankings			
Rank	State	Top rate (2017, %)	State	Top rate (2001, %)	Top Rate (2017, %)	Change (pp)
			Treatment (Group T)			
1	California	12.30	California	9.30	13.30	4
2	Oregon	9.90	New Jersey	6.37	8.97	2.6
3	Minnesota	9.85	Connecticut	4.50	6.99	2.49
4	Iowa	8.98	Minnesota	7.85	9.85	2
5	New Jersey	8.97	New York	6.85	8.82	1.97
	Average	10.00	Average	6.97	9.39	2.41
			Control (Group C)			
47	Texas	0.00	Utah	7.00	5.00	-2
48	Florida	0.00	North Carolina	7.75	5.50	-2.251
49	Washington	0.00	New Mexico	8.20	4.90	-3.3
50	Tennessee	0.00	Montana	11.00	6.90	-4.1
51	Nevada	0.00	North Dakota	12.00	2.90	-9.1
	Average	0.00	Average	9.19	5.04	-4.15

b) Use IRS state level data in excel format for tax year 2017 at (link here) to compare the fraction of high income earners in states in group C and states in group T. Fraction high earners is defined as the ratio of number of tax returns with AGI above $1m to all tax returns in group.

Under what assumption does this comparison identify the effects of state income tax rates on mobility? Is this assumption realistic (how could it be tested)?

If this assumption holds, what is the elasticity of the number of high earners with respect to the net-of-tax rate at the state level?

Group C	Fraction (%) (High Earners)	Group T	Fraction (%) (High Earners)
Texas	**0.35**	California	**0.46**
Florida	**0.42**	Oregon	**0.22**
Washington	**0.35**	Minessota	**0.27**
Tennessee	**0.22**	Iowa	**0.16**
Nevada	**0.31**	New Jersey	**0.48**
Average	**0.33**	Average	**0.32**

Assumption: Exogenous state tax rates. Unlikely to be realistic

If the assumption holds then the elasticity is equal to $e = \frac{dh}{d(1-\tau)} \frac{1-\tau}{h} = \frac{\frac{0.33-0.32}{0.33}}{\frac{100-90}{100}} = 0.30$, with h being the share of high earners by state and τ is the top marginal tax rate.

c) TCJA (the Trump tax cut) imposed a cap of $10K on state and local income taxes that taxpayers can deduct in their itemized deductions. This implies for high earners, state income taxes are no longer deductible. Explain how this magnifies the impact of state income taxes on the net-of-tax (one minus the marginal tax rate) when combining both federal and state income taxes.

Before TCJA: with deductibility, the net-of-tax rate is $(1 - \tau_{fed})(1 - \tau_{state}) = 1 - \tau_{fed} - \tau_{state} \cdot (1 - \tau_{fed})$

After TCJA: with no deductibility, the net-of-tax rate is $1 - \tau_{fed} - \tau_{state}$

So the net-of-tax rate increases by $\tau_{state} \cdot \tau_{fed}$ due to TCJA

d) Use IRS state level data in excel format for tax years 2017 and 2018 at (link here) to compare the changes in the fraction of high income earners in states in group T and states in group C from 2017 to 2018. Fraction high earners is again defined as the ratio of tax returns with AGI above $1m to all tax returns.

Construct the DD estimate using the variation created by TCJA that was discussed in c). What is the elasticity of the number of high earners with respect to the net-of-tax rate at the state level that you obtain?

Do you find this estimate more compelling that the one obtained in question a)? Why or why not?

The table gives the fraction of high earners in 2018,

Group C	Fraction (%) (High Earners)	Group T	Fraction (%) (High Earners)
Texas	**0.36**	California	**0.50**
Florida	**0.43**	Oregon	**0.24**
Washington	**0.43**	Minessota	**0.29**
Tennessee	**0.26**	Iowa	**0.18**
Nevada	**0.36**	New Jersey	**0.5**
Average	**0.37**	Average	**0.34**

Changes in the fraction of high earners in groups T and C are 0.027% and 0.036%, respectively. Using the net-of-tax rate variation calculated in c) we find the elasticity e_{DD}=0.5. This approach is more reliable because it partials out time trends.

2. Bunching at kink points (3pts)

a)

$$\max wh - T(wh) - \frac{h^{1+k}}{1+k}$$

FOC h: $w(1-T') = h^k$ hence $h = w^{1/k}(1-T')^{1/k}$ and $z = wh = w^{1+1/k}(1-T')^{1/k}$:
Hence, three cases depending on size of w:
+ if $w \leq \bar{z}^{k/(k+1)}$ then $z = w^{1+1/k}$. This is the first bracket.
+ if $\bar{z}^{k/(k+1)} \leq w \leq \bar{z}^{k/(k+1)}/(1-\tau)^{1/(k+1)}$ then $z = \bar{z}$. This is bunching at $\bar{z}$.
+ if $\bar{z}^{k/(k+1)}/(1-\tau)^{1/(k+1)} \leq w$ then $z = w^{1+1/k}(1-\tau)^{1/k}$. This is the second bracket.

b) Elasticity is $1/k$.
Fraction bunching is $\int_{w_1}^{w_2} f(w)dw$ where $w_1 = \bar{z}^{k/(k+1)}$ and $w_2 = \bar{z}^{k/(k+1)}/(1-\tau)^{1/(k+1)}$

c) Histogram attached (created with matlab).
Histogram shows bunching at \$10,000 which is $\bar{z}$.

d) All individuals with w in (w_1, w_2) bunch at $\bar{z}$.
Absent the tax rate τ, those with wage w_1 would earn $\bar{z}$ and those with wage w_2 would earn $w_2^{1+1/k} = \bar{z}/(1-\tau)^{1/k}$.

Excess bunching is 193 individuals (with earnings exactly equal to \$10,000). There are also 193 individuals on the left of the kink with earnings between \$10,000-\$827 and \$10,000-\$1. So, absent the kink, those bunching taxpayers would have spread across a band of width \$827 approximately.

Hence \$827= $\bar{z}[1 - 1/(1-\tau)^{1/k}]$.

which translates into $e = 1/k = \log(1 - 827/10000)/\log(1 - 0.3) = 0.24$ which is very close to the 0.25 I have used to simulate the data.