

Tax Revenue Cyclicalilty and Income Inequality

Summary of Preliminary Results

Yiyang Chen

February 2022

I've compiled personal income tax data in 25,401 ZIP code areas in the United States during 2005-2012. The data is collected from Internal Revenue Service (IRS) "[SOI Tax Stats - Individual Income Tax Statistics - ZIP Code Data \(SOI\)](#)". These ZIP code areas account for around 60% of all ZIP code areas in the U.S., while the rest of ZIP code areas do not make consistent appearance in the IRS data.

I calculated approximate Gini coefficient for each ZIP code areas by using the total tax collected and total number of exemptions (which approximates the population) in each AGI class. I'm going to verify this method by applying it to the county level data and comparing against the available official statistics.

In this summary of preliminary results, I will give

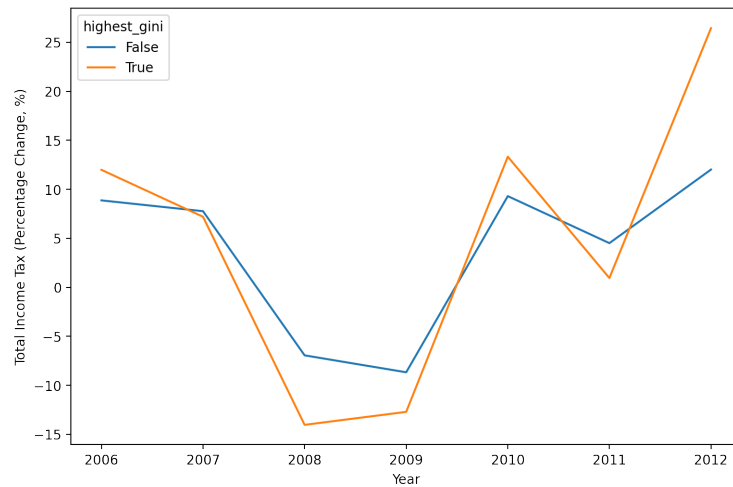
1. Summary statistics table
2. Time series plot
3. Regression result (basic setup)

Table 1: Summary Statistics

	n1	n2	AGI	Total Tax	Change	Gini
count	175,287	175,287	175,287	175,287	175,287	175,287
mean	5,462.3	10,875.3	319,978.6	41,636.8	4.8619	0.5755
std	6,595.0	13,285.2	500,672.3	89,709.1	36.970	0.0860
min	45	92	2,708.0	97.0	-93.0652	-0.3975
median	2,381	4,805	111,402.0	10819.0	4.3681	0.5709
max	98,117	135,791	18,706,420.0	4,490,904.0	131.7301	0.9560

Notes:**n1:** Number of returns, which approximates the number of households**n2:** Number of personal exemptions, which approximates the population**AGI:** Adjusted gross income**Total Tax:** Total personal income tax revenue**Change:** Percentage change in total personal income tax revenue**Gini:** Approximate Gini-coefficient calculated using AGI classes

Figure 1: Total Personal Income Tax in Highest 1% Gini Coefficient ZIP Code Areas vs All Others



In Figure 1, I take the ZIP code areas with the highest Gini coefficient and compare them against the rest. It is pretty clear that in places with higher income inequality, the personal income tax revenue responds more to the business cycles. During the GFC, ZIP code areas with the highest income inequality suffered on average around 5% more decline in personal income tax revenue.

Table 2: Dependent Variable: Percentage Change in Income Tax Revenue

Variable	Simple Model
const	1.4714*** (0.0803)
rgdp pct chg	2.6642*** (0.0331)
rgdp pct chg * highest gini	1.6109*** (0.4234)
Time fixed effect	No
ZIP code fixed effect	No
R-Sqaure	
Overall	0.0168
Between	0.8387
Within	9.164e-05
N	175,284
F-statistics	3252.0
p-value	0.0000

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2 reports the preliminary regression results. For every 1% increase in the real GDP, the personal income tax revenue is expected to increase by 2.66%. For those ZIP code areas with highest income inequality, the change in tax revenue is about 4.27% per 1% change in real GDP.

Also, even though the overall R-square is low, the model gives a very high Between R-square. This indicates the model is capable of capturing the variation in the change in tax revenue between the ZIP code areas.