

# The Finnish Top 0.4 Percent: An Exploration of Top Tax Shares in Finland from 2009 to 2013

*Kyle Ott and Cornelius Schneider*

*14 November 2014*

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Our Unique, Tidy, Open, Reproducible Dataset</b>	<b>2</b>
<b>3</b>	<b>Descriptive Statistics of Our Data</b>	<b>3</b>
<b>4</b>	<b>Inferential Statistics: First Model Attempts</b>	<b>10</b>
<b>5</b>	<b>Conclusion: Further Analysis Needed</b>	<b>12</b>
	<b>References</b>	<b>13</b>

## 1 Introduction

The question about an “optimal” income taxation was always discussed against the background of classic economic theories: Income taxation should maximize a given social welfare function, that depicts a societies preference for equality. Furthermore, sacrifice theory of income taxation illustrates that redistribution should take place up to the point where marginal utilities are equalized.

However, these theories completely neglect behavioral responses to taxation. According to the Laffer-Curve at one point a further increase of the tax rate would result in a loss in tax revenues due to negative labor supply responses. In a very recent work, (Piketty and Saez 2014) identifies additional reasons, why trends in top-income shares are correlated with the tax rates: Labor responses, evasion/avoidance responses and bargaining responses.

Whereas poverty is studied extensively in Economics through surveys and welfare programs, the current debate still lacks information about the top of the income distribution. The aim of this project is to inspect exactly this upper end of the income distribution. For this, we'll analyze micro-level data from finish taxpayers of the years 2009 to 2013.

In Finland the tax on earned income is levied according to a progressive tax scale: Each tax payer has to pay a basic amount dependent on his earned income plus the tax rate within the respective tax bracket. The concrete tax scheme is decided annually by the Parliament. The relevant tax rates in the top tax-bracket were levied as follows:

- 2009: 30.5%
- 2010: 30.0%
- 2011: 30.0%
- 2012: 29.75%
- 2013: 31.75%

Against this background, our paper's purpose is to identify anomalies in the tax patterns of ultra-wealthy finish people. In a first step, we will visualize how the share of the top 0.4% (wealthiest 15.000 individuals) in total income-tax revenue changes between 2009 and 2013. Following up, with inferential statistics we try to dig deeper into income-tax payer's behavior: What is the predicted probability to pay a certain average tax rate given your income? What was the potential impact of changes in top tax-rates on ultra-wealthy finish people? Conditional on the first results, more precise models, as well as a suitable strategy to tackle the raised questions will be assessed.

## 2 Our Unique, Tidy, Open, Reproducible Dataset

The Nordic countries Finland, Sweden and Norway have a tradition of publishing everyone's income and tax details every year. Whereas in Sweden and Norway this data is only accessible being a citizen and after pulling an official request, in Finland top income tax earners are public figures as a result of heightened media scrutiny on top income tax earners. Finland's largest business online daily newspaper “Taloussanomat” published the figures in a suitable format for our purpose: The top 15.000 income earners are displayed on yearly basis from 2009 to 2013, including their total income, taxed paid and average tax rate.

In a first step we scraped the data from their homepage. Given that each year's data was partitioned into 30 tables of 500 observations each, we wrote a loop in order to gather the date in one data frame. We ran that loop for each year and appended them into the same data frame. Moreover, we converted the strings with weird characters and various lengths into numeric class. Since there are some observations missing in the first three years (2009 to 2011), the final data set contains 70.402 observations.

Afterwards, we merged that data frame with two tables from the OECD: One indicating the total Finnish income-tax revenues from 2009 to 2013 and one indicating the number of working population. In addition to

that, we manually amended the total GDP in constant prices (national base year) as well as the respective top-tax rates for each year.

Based on this data, we created a couple of new variables that are helpful for our analysis: The rank based on the income of each individual in every year; the change in the tax rate compared to the previous year and the share of total income-revenues that the top 15.000 tax payers pay.

Before we continue to the descriptive analysis, just one note on the origins of our various data. First, as already mentioned, all of the micro-level income data comes from taloussanomat, the income tax revenue for 2013 comes from Finland, the income tax revenue for 2009-2012 comes from Stats, the population data comes from Stats, the income tax scheme for 2013 comes from eTax, the income tax scheme for 2012 comes from KPMG, the income tax scheme for 2011 comes from Wikipedia, our income tax scheme for 2010 comes from International (2010), and the income tax scheme for 2009 comes from Finance.

### 3 Descriptive Statistics of Our Data

In the table below, you will see the breakdown of the data by: year, number of observation (“n”), total working population in Finland in given year, share in working population presented by top 15,000 income earners, total tax revenues collected (in Euros), total GDP (in constant prices national base year), top income tax rate, and the change of this tax rate by year.

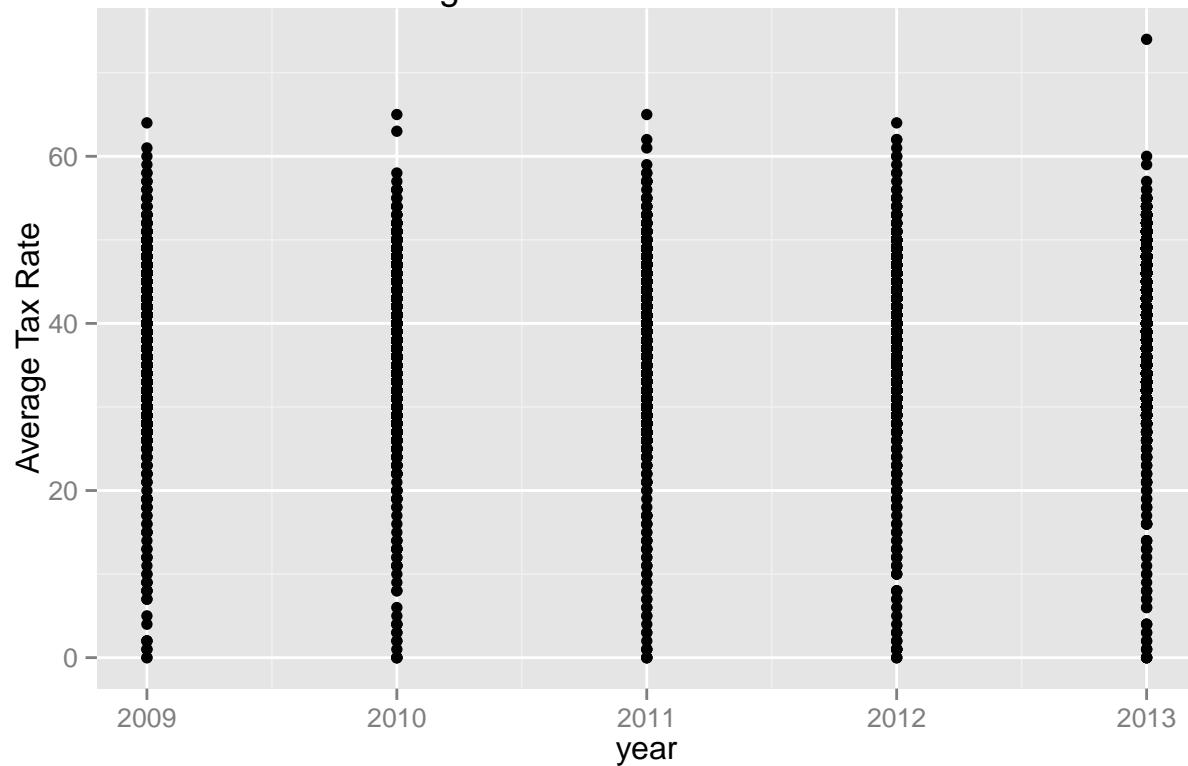
year	n	WorkPop	percent	Total_Tax_Revenue	Total_GDP	Tax_Rate	DELTA_Tax_Rate
2009	12134	3547335	0.34	2.650e+10	1.817e+11	30.50	NA
2010	14290	3549338	0.40	2.710e+10	1.871e+11	30.00	0.50
2011	13978	3538686	0.40	2.933e+10	1.919e+11	30.00	0.00
2012	15000	3524349	0.43	2.920e+10	1.891e+11	29.75	-0.25
2013	15000	3508645	0.43	3.078e+10	1.868e+11	31.75	1.00

The table below shows a quick summary statistics about the distribution of our data and three key variables of interest over all five years: total annual income, total paid taxes, and the average tax rate.

table3	tot_inc	tot_paid_taxes	ave_tax_rate
Median	207456	84494	43.0
Mean	315287	119433	40.3
SD	1067760	340304	7.2

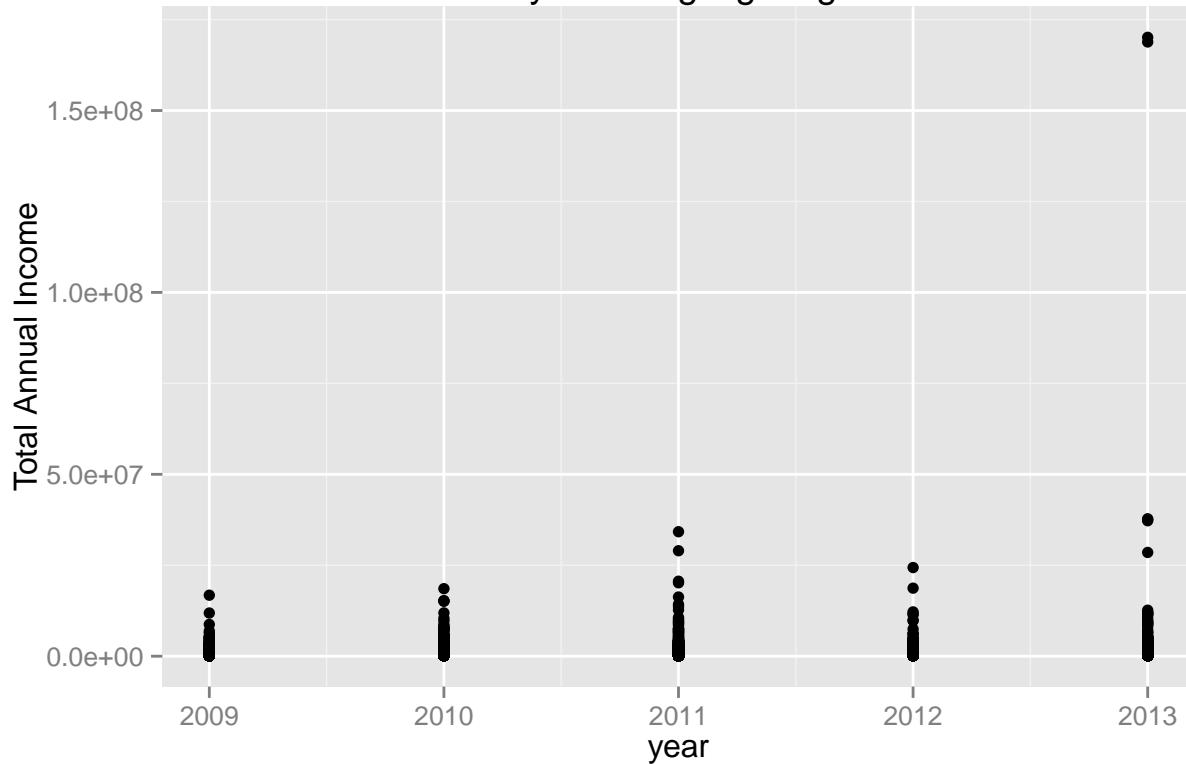
The figure below shows the range of average tax rates paid each year across all observations. It was included to visually display the range as it slightly changes across time.

### Average Tax Rate Paid Across Years



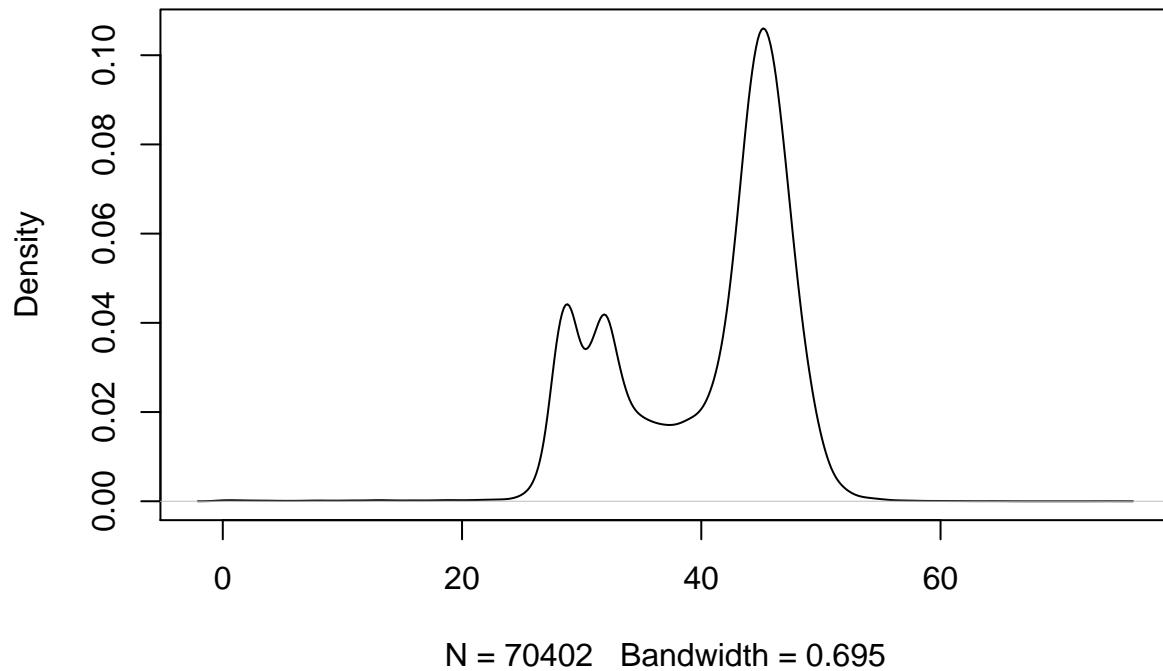
The next plot was included to show the total income outliers that are very clear in 2013 (top right corner). In future analysis, we will consider dropping these extreme outliers (recent technology millionaires).

## Total Income by Year Highlighting 2013 Outliers



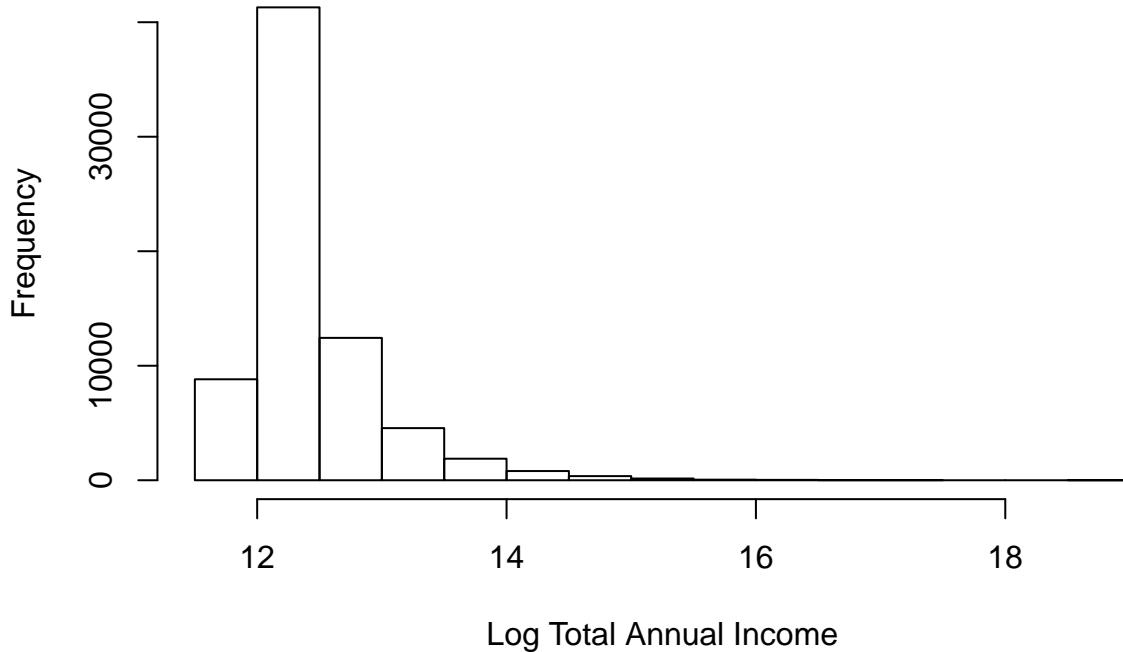
The next plot shows the density plot of the average tax rates for all years. It is interesting to note that there is a bimodal distribution. We cannot make any inferences about why this may be the case, because we are not able to observe actual behavoir responses due to data limitations. Moreover, almost all of the observations fall between roughly 25 and 55 percent.

## Density Plot of Average Tax Rates



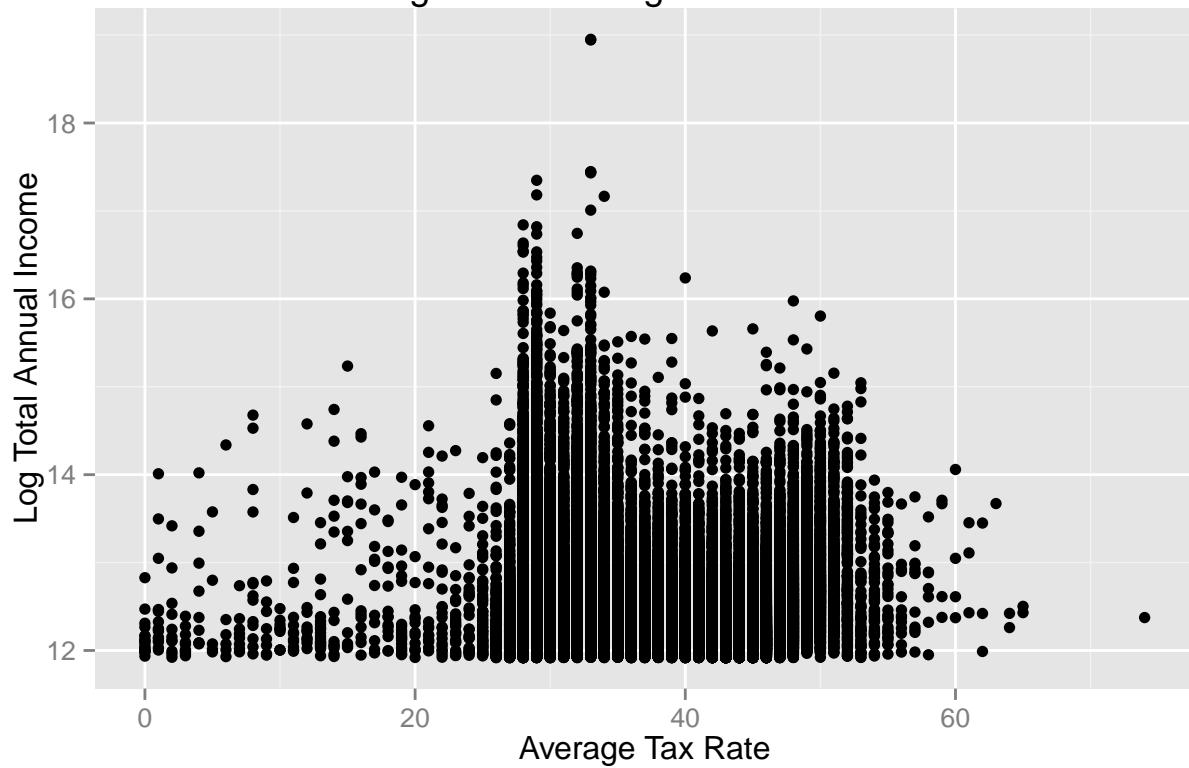
The plot below shows the histogram of the total annual income after it has been logged. Before it was lot it was extremely far from normal, now it is a bit closer but still not ideal. However, we are aware that our data suffers from skew because we are focusing only on a specific subsample, namely the ultra wealthy.

## Histogram of Logged Total Income



The next figure plots the log total annual income against the average tax rate. Here, in our visual outlook of our data, we are interested in observing patterns between these two key variables. There are clearly multiple instances where higher income individuals have a small average tax rate, which is likely due to potential tax avoidance like for example, shifting income to capital gains. This figure is one in which we would like to develop further for our final project.

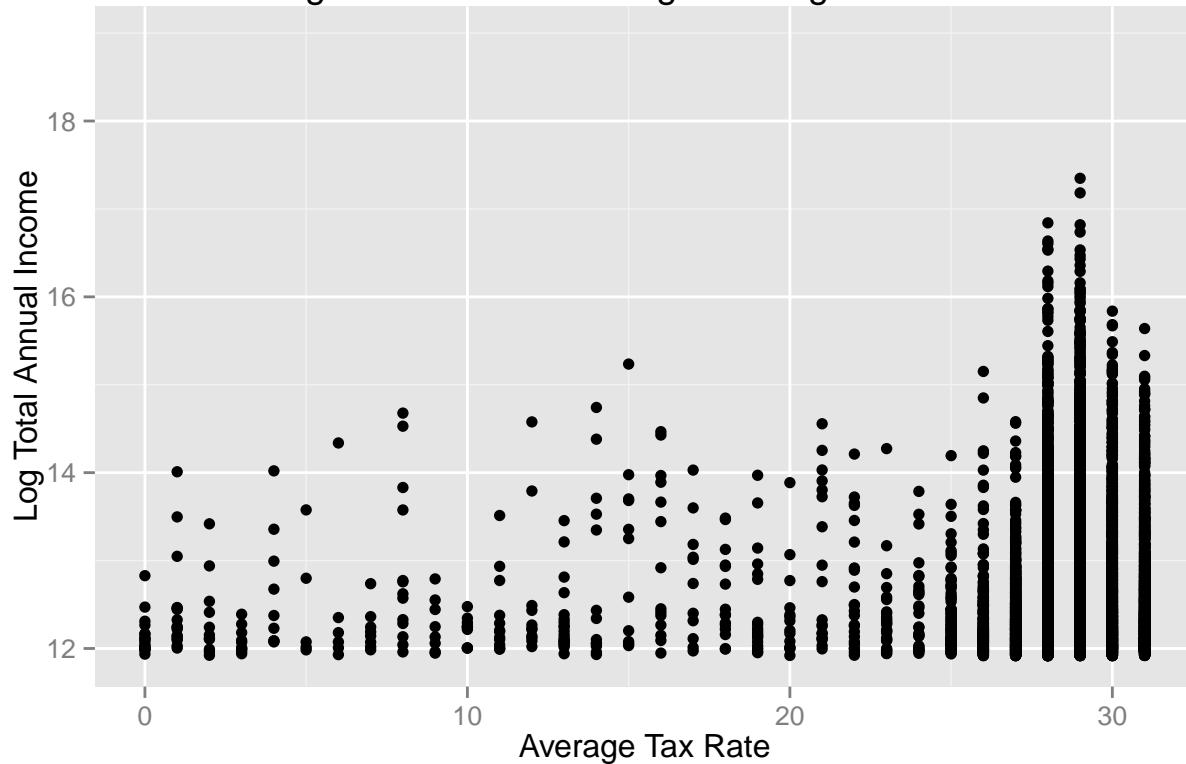
Plot of Average Tax Rate Against Annual Total Income



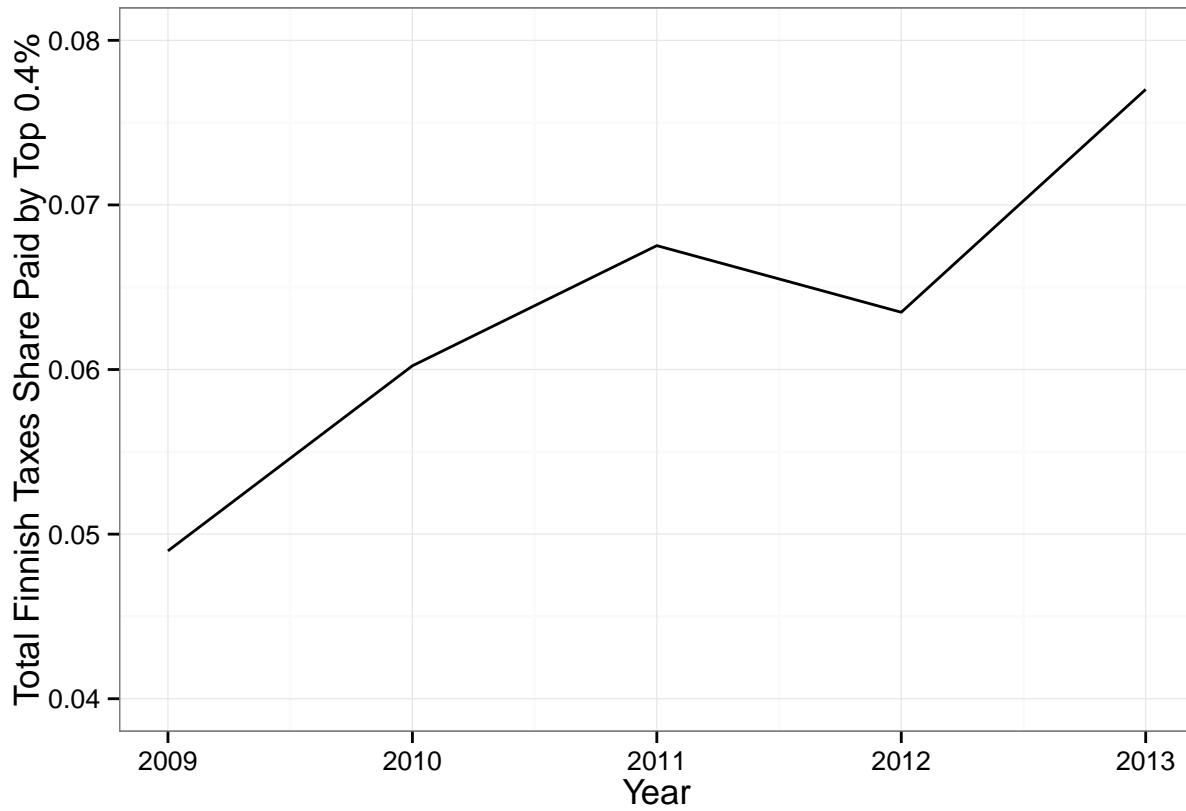
The following plot is a zoomed in version of the previous plot. It shows the frequency of tax payers who are named above and likely engaging in tax avoidance schemes and resulting with low average tax rates.

```
## Warning: Removed 57870 rows containing missing values (geom_point).
```

Plot of Average Tax Rate < 31% Against Log Annual Total Incomes

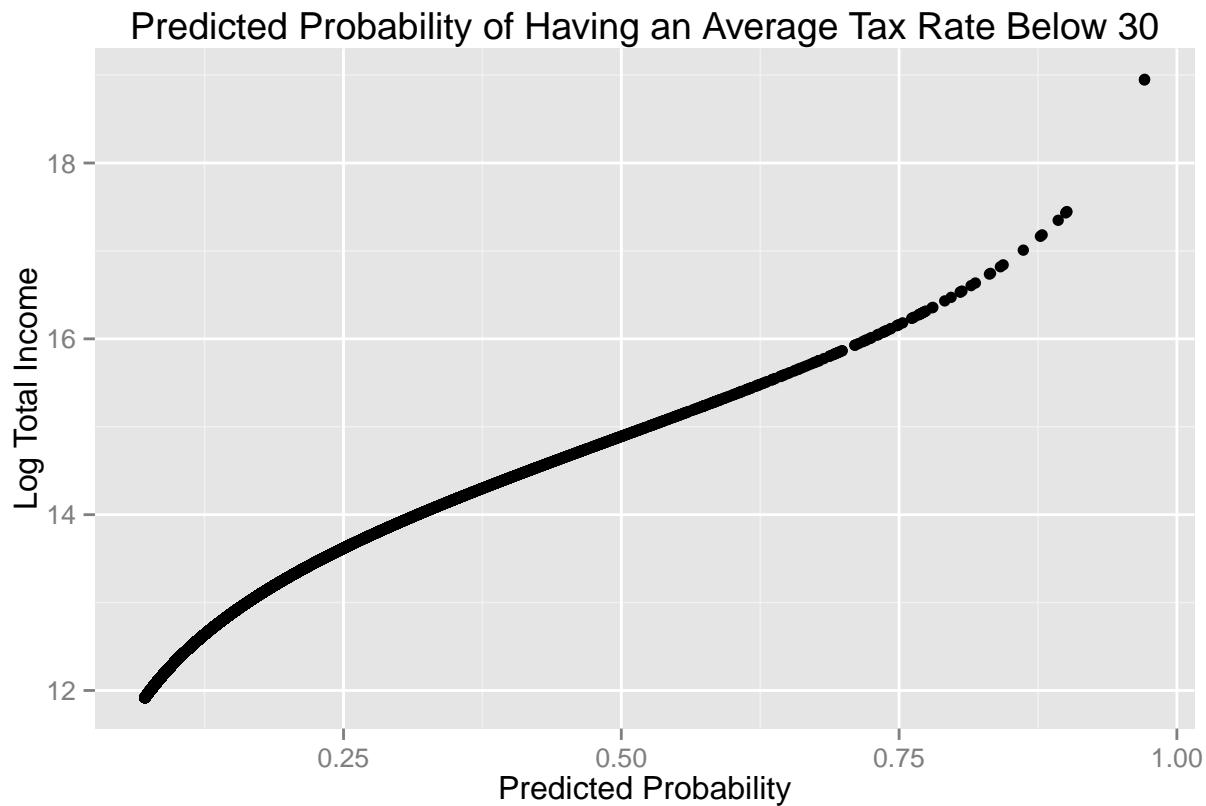


The line graph depicts the development of the share of the total income tax revenue paid by the top 0.4 percent, which is our sample. Over the course of five years, the share increase from 5 percent to almost 8 percent. An in each year, we can nicely observe the changes in the top income tax rate changes which likely resulted in the reflected changed in tax revenue collected by the top 0.4 percent.



## 4 Inferential Statistics: First Model Attempts

Our first attempt at running a model is extremely simple. Essentially, we want to see and develop in further analysis a relationship between a given income and the predicted probability of paying an average tax rate below certain thresholds. In our first attempt, we ran a logit model with a binary dependent variable (with 1 being observations with a average tax rate below 30). The results are best illustrated by the following figure, a predicted probability plot. The curve of the probabilities is intriguing. The maximum income has the highest probability of paying a tax rate below 30 percent. This is a strong result. Moreover, the curve suggests that lower incomes (within our sample) have a lower probability of paying an advantageous average tax rate. This finding can be seen as motivation for further inspection and analysis.



The (simple) “dynamic” regression output below is our failed attempt to run a legitimate and appropriate OLS model including our macro data. Going forward, we hope to find a way to run an model that is able to overcome our initial difficulties.

```
##
## Call:
## lm(formula = share ~ Tax_Rate + ratio + DELTA_Tax_Rate + log(Total_GDP),
##      data = FINAL)
##
## Residuals:
##      Min        1Q        Median        3Q        Max 
## -4.5e-10  0.0e+00  0.0e+00  0.0e+00  1.0e-13 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -4.89e+00  2.86e-11 -1.71e+11 <2e-16 ***
## Tax_Rate     1.13e-02  2.11e-14  5.37e+11 <2e-16 ***
## ratio        1.52e-15  1.09e-15  1.40e+00   0.16    
## DELTA_Tax_Rate -5.58e-03  4.30e-14 -1.30e+11 <2e-16 ***
## log(Total_GDP) 1.78e-01  1.11e-12  1.60e+11 <2e-16 ***  
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.86e-12 on 58263 degrees of freedom
##   (12134 observations deleted due to missingness)
## Multiple R-squared:      1,   Adjusted R-squared:      1
```

```
## F-statistic: 1.69e+23 on 4 and 58263 DF, p-value: <2e-16
```

Table 3: Inappropriate Model, First Try

<i>Dependent variable:</i>	
	share
(Intercept)	0.011*** (0.000)
Top tax rate	0.000 (0.000)
Change in top tax rate	-0.006*** (0.000)
Log total GDP	0.178*** (0.000)
Constant	-4.885*** (0.000)
Observations	58,268
R <sup>2</sup>	1.000
Adjusted R <sup>2</sup>	1.000
Residual Std. Error	0.000 (df = 58263)
F Statistic	168,984,712,956,700,518,252,544.000*** (df = 4; 58263)

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## 5 Conclusion: Further Analysis Needed

Our attempt to analyze the top income-tax share dependent on the tax rate by running an ordinary least squares model has failed so far. Obviously controlling for GDP and considering year dummies didn't lead to the expected outcome: We're still considering to incorporate other data to build a worthy model. Finally, perhaps our focus should shift to expanding on our simple logit model and incorporating other variables and also tweaking our depending variable looking at different thresholds and also different years.

---

This project used RStudio (2014) to create this assignment.

## References

- eTax, Nordisk. “Tax Rates Finland.” <https://www.nordisketax.net/main.asp?url=files/suo/eng/i07.asp>.
- Finance, Ministry of. “Taxation in Finland 2009.” [https://www.vero.fi/download/Taxation\\_in\\_Finland\\_2009/%7BF740042A-A01D-49ED-8B3F-FBD3420D19D7%7D/7714](https://www.vero.fi/download/Taxation_in_Finland_2009/%7BF740042A-A01D-49ED-8B3F-FBD3420D19D7%7D/7714).
- Finland, Statistics. “Taxes by Sector and Tax Category.” [http://www.stat.fi/til/vermak/2013/vermak\\_2013\\_2014-07-11\\_tau\\_001\\_en.html](http://www.stat.fi/til/vermak/2013/vermak_2013_2014-07-11_tau_001_en.html).
- International, Crowe Horwath. 2010. *Worldwide Business Tax Facts 2010/11*. CCH Australia Limited.
- KPMG. “Tax Rates/Brackets.” <http://www.kpmg.com/US/en/IssuesAndInsights/ArticlesPublications/flash-international-executive-alert/Documents/flash-international-executive-alert-2012-018-jan.pdf>.
- Piketty, Thomas, and Emmanuel Saez. 2014. “Inequality in the Long Run.” *Science* 344 (6186). sciencemag.org: 838–43.
- RStudio. 2014. *RStudio*. Boston, MA. <http://www.rstudio.com/>.
- Stats, OECD. “Income Tax Revenue 2009-2012.” <http://stats.oecd.org/restsdmx/sdmx.ashx/GetData/REV/NES.1000.TAXNAT.FIN?startTime=2009&endTime=2012>.
- . “Population 2009-2013.” [http://stats.oecd.org/restsdmx/sdmx.ashx/GetData/POP\\_FIVE\\_HIST/FIN.YP99TLL1\\_ST.TT.A?startTime=2009&endTime=2013](http://stats.oecd.org/restsdmx/sdmx.ashx/GetData/POP_FIVE_HIST/FIN.YP99TLL1_ST.TT.A?startTime=2009&endTime=2013).
- Wikipedia. “Taxation in Finland.” [http://en.wikipedia.org/wiki/Taxation\\_in\\_Finland](http://en.wikipedia.org/wiki/Taxation_in_Finland).