

Public's Tax Preference *

Keertana Chidambaram[†]

June 2019

(version 1.05)

Abstract

The two key research questions that this study is aiming to answer are: (1) do people favor tax structures that financially benefit them directly? And (2) how regressive or progressive are these tax preferences? To answer these questions, this study administers a survey to gather data from a representative sample of the population. It is found that compared to the current policy, 28% prefer increase in tax revenue, 32.7% are willing to pay more income tax, 34.5% and 19.5% prefer a more progressive income and long-term capital gains tax respectively. The collected data verifies the established claim that Republicans on average prefer lower and a more regressive tax structure. But it is found that only young Republicans exhibit this general trend. On the other extreme, old Republicans prefer the same progressive and pro-tax policies as old Democrats. Between men and women who aren't primary earners women prefer more taxes. Finally, it was found that random forests were better predictors of tax behaviours from demographic variables than decision trees.

keywords: Tax preference, progressive taxation.

JEL classification: C8, H2

*I would like to thank the Open Source Economics Lab at the University of Chicago for funding this study. I would also like to thank Dr. Richard Evans for his continuous support and guidance.

[†]<https://github.com/keertanavc>, University of Chicago, MA Program in Computational Social Science, Department of Social Sciences, 5730 S. Woodlawn Avenue, Chicago (315) 866-5195, keertana@uchicago.edu.

1 Introduction

Taxation has become a frequently discussed issue in the USA following the tax-cut reforms of 2018 [Cole (2015), Nunns et al. (2016)] and more recently with the Green New Deal wave, calling for marginal tax rates as high as 70% for the top earners in the country [Saez and Zucman (2019), de Rugy and Salmon (2019)]. With this growing interest in taxes, several organizations have reported on the public's opinion on taxes [Pew Research Center (2018), Pew Research Center (2019), Murray (2019)]. This study focuses on two aspect of tax opinion: self-interest and progressivity. More specifically the two research questions askeed are: (1) do people favor tax structures that financially benefit them directly? And (2) how regressive or progressive are these tax preferences?

Several studies have explored how self-interest drives tax preferences. Beck et al. (1987) finds that people's tax preferences are motivated by self-interest. In the study, most Florida residents preferred raising taxes that do not affect them such as taxing new homeowners in the area. In Ballard-Rosa et al. (2017) also the respondents exhibited self-interest. In the survey, on average, respondent's preferred tax structures assigned lesser rates on their own income bracket.

On the other hand, the public in general demands more services from the government. Citrin (1979) finds that most respondents have the opinion that they pay higher than average tax. But concurrently, they also demand more spending and services from the government. The anti-tax sentiments varied with respect to three factors. First, imparting information about government cutting when taxes are low decreased the sentiment. Second, the sentiment was stronger when direct taxes were raised as compared to invisible taxes. Another study by the Pew Research Center (2011) reveals that about half the American population feels that they are paying the right amount of taxes. But a majority (60%) feel dissatisfied with the current tax structure. Beck et al. (1987) find that a majority (50%) find taxes to be too high but are satisfied with public services. The majority also preferred paying more taxes over sacrificing public goods.

Another interesting fact that needs to be given adequate consideration is the pro-tax behaviour exhibited by certain individuals. Several billionaires including Warren Buffet, Howard Schultz, and Bill Gates have openly aired their opinion in the media supporting a higher tax on their income and wealth. [World Bank \(2019\)](#) finds that in Brazil, people are more willing to pay more taxes if the government is transparent about its spending. Also, Americans feel strongly about the ethics of paying due taxes, according to [World Bank \(2017\)](#), "The majority of Americans (88%) say it is not at all acceptable to cheat on taxes". Thus there is conflicting evidence about the tax behaviour of individuals.

Growing inequality has sparked numerous debates about what a more equitable tax structure should look like. So, quite unsurprisingly, several studies survey the support for progressive tax structures. [Ballard-Rosa et al. \(2017\)](#) finds interesting trends in America's support for progressive taxation. There was strong disapproval for tax plans that taxed the poor heavily. But although on average respondents approved taxing the rich more, there was no significant difference in support between tax brackets that assigned values ranging from 25% to 50% for the rich. So, while there was a majority agreement on the existence of inequality, people preferred less tax on the poor but flat tax on the rich. Overall most people preferred tax structures that raised less total revenue than the existing tax structure. A treatment group was 'nudged' by being shown a rough estimate of the dip in revenue from the new tax plan. These respondents preferred structures that raised more revenue than the control group but even their preferred tax structure raised less revenue than the existing structure. [Roberts et al. \(1994\)](#) finds that most people support the idea of a progressive tax structure. But when presented with tax brackets and rates, most people pick a flat/regressive tax structure.

[Gaines \(2017\)](#) compiles the results of a group of surveys about America's public opinion on taxes. One study finds that although many Americans believe the rich do not pay enough taxes, they estimate the 'fair' tax for the rich to be lower than the current tax rates. In another study, respondents were asked to pick a tax structure, and then they were shown the rough estimate for the revenue raised. Most

respondents who picked a tax structure with lower revenues voted to cut government spending followed by taxing the rich to make up for the shortage. When forced to pick a tax structure that raised a minimum revenue, the average respondent raised the taxes everyone except for the poor population. Another study found that an average American felt that a fair lottery tax should be 15%, whereas it was 35% not including state taxes. These studies also reinforce the fact that most respondents do not keep government revenues in mind while choosing their favorable tax. It also translates that general knowledge about the tax rates for different brackets and progressive taxation is lacking.

Studies also outline the role of demographics in tax preference. [Citrin \(1979\)](#) found that tax preferences varied over various demographics. White, homeowners, and Republicans were more likely to support tax cuts than black, renters, and Democrats. [McGee et al. \(2019\)](#) show that several factors like age, gender, education, income, religion, country, etc. played a role in determining an individual's tax preference. [Ballard-Rosa et al. \(2017\)](#) show that democrats and republicans both on average prefer a progressive structure, but Republicans preferred an overall less tax. Republicans support a high tax on the rich, but this support was lower than the support from Democrats.

How is the questions answered: data, theory, methods, whats' new Most previous surveys estimate the effect of change in tax on the macro scale (i.e. on the income of different groups and total revenue) based on some very approximate models. The Open Source Tax Calculator [[Tax-Calculator 2.2.0 \(2019\)](#)] is used for the macro-level simulation of the effects of tax changes in this study. Most previous studies also measure tax progressivity as a simple run-over-rise measure. But this become more and more erraneous as the number of tax brackets increases. In the current US tax structure, there are 7 brackets for personal income. To rectify this problem, the Suit's index is used to measure progressivity. This study also thoroughly investigates the effect of demographic characteristics on tax preferences. Further, tax behaviour is predicted using the two machine learning models decision trees and random forests.

Compared to the current tax policy, 28% support increase in tax revenue, 32.7%

are willing to pay more income tax, 34.5% and 19.5% prefer a more progressive income and long-term capital gains tax respectively. The claim that Republicans prefer lower and more regressive taxes is validated. But there is a significant age effect on this trend with young republicans displaying the same traits (regressive and lower taxes) as the group as a whole but old republicans and older other party members prefer the same progressive and pro-tax policies. Between men and women, if they are primary earners in the household, there is no difference in preferred policies but between those who are not primary earners, women exhibit more pro-tax behaviour by preferring more taxes and lower after-tax incomes. It is also found that random forests predict tax preferences from demographic variables with higher predictive accuracy than decision trees.

The rest of the paper is structured as follows: first the survey design is explained in detail, then the methodology used to calculate the effect of preferred tax-variables is explained. This is followed by a data section summarizing the collected responses. Then the analysis and results section details the trend analysis and model fitting performed on the data and the corresponding results and inferences. Finally the conclusions section summarizes the results and charts out the limitations and future scope for the study.

2 Methodology

Survey Design

Figure 1 shows how the survey is structured. First participants are shown an informational video about tax structures and marginal tax rates. This is important because according to the existing literature, general knowledge about how taxes work is lacking [Gaines (2017)]. Then, all participants fill out two separate surveys: one for their demographic information and the other for their tax preference¹. The demographic survey captures the nine variables: gender, age, political affiliation, his/her

¹Check out [this link](#) for the full survey

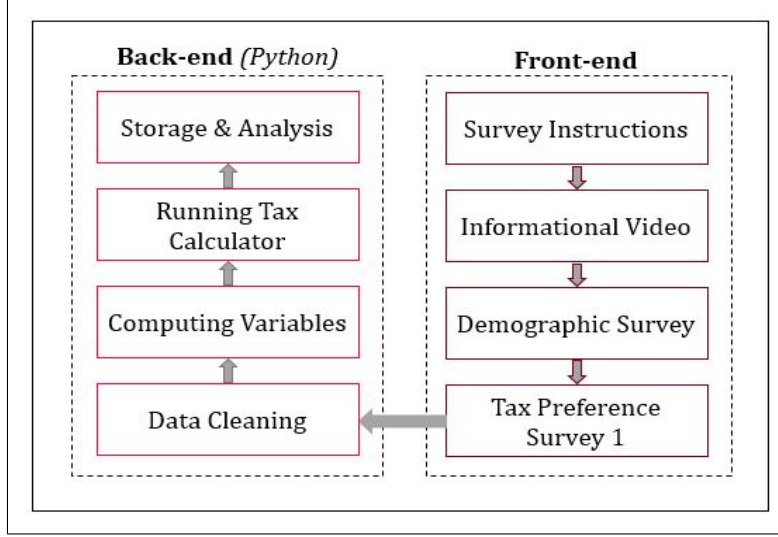


Figure 1: Survey organization

level of education, race, annual household income for 2018, marital status, number of children, and if the respondent is the primary earner of the household or not. The tax preference survey captures the eleven variables: the maximum taxable earnings for social security, the long-term capital gains for all three brackets, and the personal income tax for all seven brackets. The range of values these variables can take along with the current values of the variables and bracket information is given in Table 5 in the Appendix section of this paper. Figure 2 shows a screenshot of a part of the survey. The survey is hosted on Amazon’s Mechanical Turk (mTurk) platform. mTurk is an online crowdsourcing platform where workers perform tasks like surveys and experiments virtually [Horton et al. (2011), Paolacci et al. (2010)].

Composite Variable Calculation

Composite variables are calculated to measure the impact of the new tax policy suggested by the respondent. The four composite variables that are calculated are: (1) change in total revenue collected by the government, (2) percentage change in income for the income bracket of the respondent, (3) Suit’s index to measure progressivity of the income tax structure, and (4) progressivity of the new long-term capital gains tax structure.

Tax Preference Survey 1

In this section, you need to respond with your most preferred values for each of the tax variables. You can either enter fresh values if you prefer something other than the existing tax policy or you can enter the current values (or leave the sliders unmoved) if your preferred values are the same as the current tax policy.

Maximum taxable earnings for Social Security (current = \$128,400) (please enter a number): This represents the maximum amount of individual earnings that are subject to the payroll tax, which funds the Social Security system.

128400

Long-term capital gains and qualified dividends tax rates:

0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100

\$0 to \$37,950 bracket (current = 0%):

\$37,950 to \$418,400 bracket (current = 15%):

Figure 2: Sample screenshot of the survey

The percentage change in revenue and income are calculated using the Open Source Tax Calculator [[Tax-Calculator 2.2.0 \(2019\)](#), `taxcalc` package in Python]. Only the parameters changed by the respondents (i.e. maximum taxable social security earning, PI taxes and LTCG taxes) are changed in the policy, every other variable is left at its original value. The data used to simulate the model is the proprietary IRS’s Statistics of Income (SOI) Public Use File (PUF). The data can be purchased from IRS. The long-term capital gains progressivity index calculated as the average increase/decrease between the lowest to the highest brackets of tax.

Progressivity measure for personal income is calculated using the Suit’s index [[Suits \(1977\)](#)]. It is based on the Lorenz curve [[Gastwirth \(1971\)](#)] and the Gini index [[Gastwirth \(1972\)](#)]. It is calculated as the difference in area under a percentage accumulated tax burden vs. accumulated total income curve for the new tax w.r.t a flat tax expressed as a percentage of area under the curve for a flat tax. [Figure 3](#) illustrates the calculation of Suit’s index for the current tax structure.

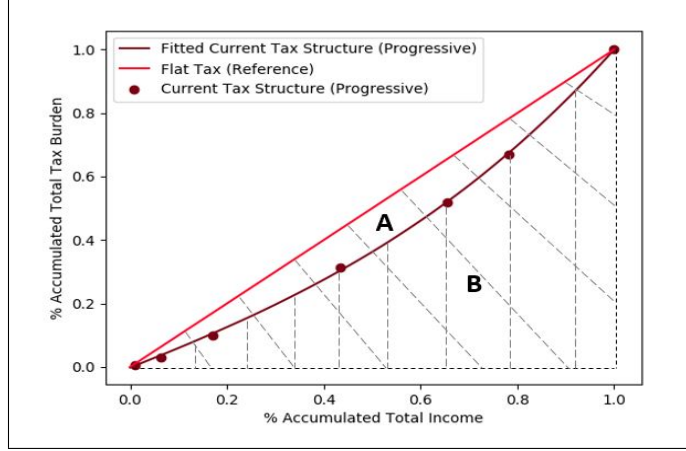


Figure 3: Suits Index Calculation for Current Tax Structure

Correspondingly, the formula is given as:

$$S_p = \frac{A - B}{A}$$

Hence the index is zero for flat, positive for progressive and negative for regressive tax structures. The aggregate values are calculated using the data in [1](#) which gives the distribution of population in each income groups. The population in each income bracket is approximated as the percentage of tax returns from that tax bracket according to the 2015 IRS income tax statistics data [[Pew Research Center \(2017\)](#)]. After obtaining the points on the graph for the given policy, a cubic spline is fit (using Python's `scipy` package [[Jones et al. \(2001-\)](#)]) to generate the curve. The area for flat tax (A) is fixed at 0.5, and the area under the progressive tax (B) is calculated by integrating the fit cubic spline. The calculated value of S_p for the current tax policy is 0.178.

Machine Learning Models

Two models are used to predict the tax behaviour of individuals using the demographic data: decision tree [[Quinlan \(1986\)](#)] and random forest [[Liaw et al. \(2002\)](#)]. A decision tree is a non-parametric supervised machine learning model that com-

S.No	Income Tax Bracket	% Population
1	\$0 to \$9,700	22%
2	\$9,701 to \$39,475	22%
3	\$39,476 to \$84,200	18%
4	\$84,201 to \$160,725	22%
5	\$160,726 to \$204,100	12%
6	\$204,101 to \$510,300	4%
7	\$510,300 +	1%

Table 1: Aggregate population and income data

binex both classification and regression. This is apt for analysis because the observed dataset has both continuous and categorical variables. Random forest consists of an ensemble of multiple decision trees fit on random samples of the data. This helps improve accuracy of prediction and controls for over-fitting as compared to the decision tree. Hence, the random forest model is expected to be a better fitting model as compared to the decision tree model.

3 Data

A total of 1080 observations were collected through the survey. Figure 4 shows the geographic locations of the respondents of the survey. The descriptive data for categorical and continuous variables collected are given by Figure 5 and Table 2 respectively. Clearly, the collected data forms a fairly representative and diverse sample.



Figure 4: Geographic dispersion of the collected data

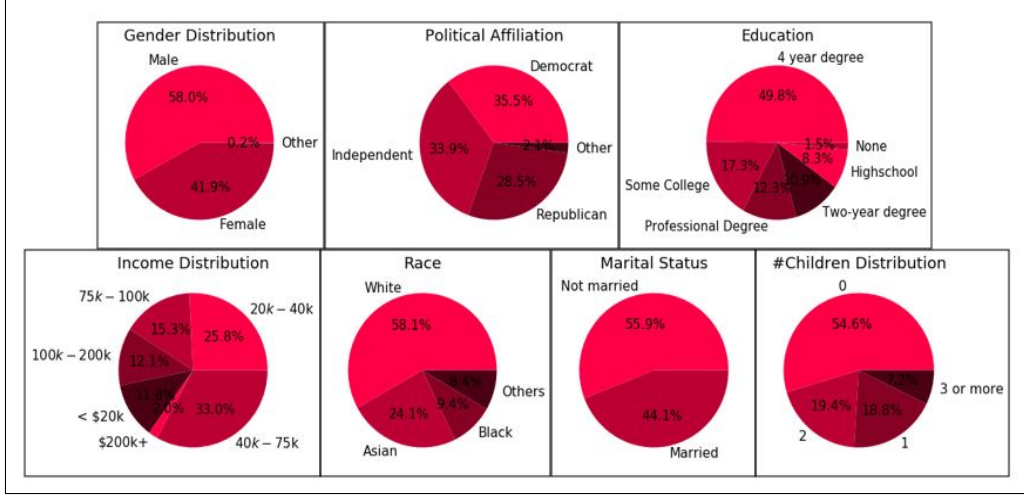


Figure 5: Composition for the discrete survey variables

4 Analysis and Results

Composite variables were computed as per the **Methodology** section of the paper. Table 3 provides descriptive statistics and Figure 6 shows the distribution for the newly calculated variables. Majority of the respondents prefer a progressive tax structure overall for both personal income and long-term capital gains. But individuals seeking lower taxes want a significant decrease, but the individuals wanting higher taxes only prefer a small increase. Compared to the current levels (i.e. % change in income = 0, change in revenue = 0, income tax progressivity = 0.178, and long-term capital gains progressivity = 0.12), 28% want an increase in revenue, 32.7% are willing to take a lower after-tax income, 34.5% opt for a more progressive income tax structure, and 19.54% want a more progressive LTCG tax structure.

To explore the effects of demographic variables on the calculated variables, 2 sample t-tests are carried out. A 2 sample t-test [Owen (1965)] is used to test if the mean of two different sample distributions are the same. This test is carried out only on distributions that have a size of at least 30 samples and effects are reported only if there is a 95% confidence, i.e. $p < 0.05$.

Effect of political affiliation: As seen in the literature review, all the previous studies support the claim that Republicans overall prefer less and comparatively more

Survey variables	Mean	5 th Percentile	95 th Percentile
Age	35	23	58
Maximum taxable earnings for Social Security	165,623	50,000	500,000
LTCG Tax 1	20%	0%	75%
LTCG Tax 2	27%	6%	76%
LTCG Tax 3	33%	10%	80%
PI Tax 1	21%	0%	76%
PI Tax 2	24%	5%	79%
PI Tax 3	30%	10%	80%
PI Tax 4	32%	12%	78%
PI Tax 5	39%	15%	81%
PI Tax 6	42%	15%	78%
PI Tax 7	46%	15%	84%

Table 2: Descriptive statistics for the continuous survey variables

Variables	Mean	Min	Max
Delta Revenue (Bil \$)	-614.07	-6607.49	604.63
Delta Income (%)	6.16	-11.94	205.93
PI Progressivity	0.16	-0.56	0.78
LTCG Progressivity	0.07	-0.5	0.50

Table 3: Descriptive statistics for computed variables

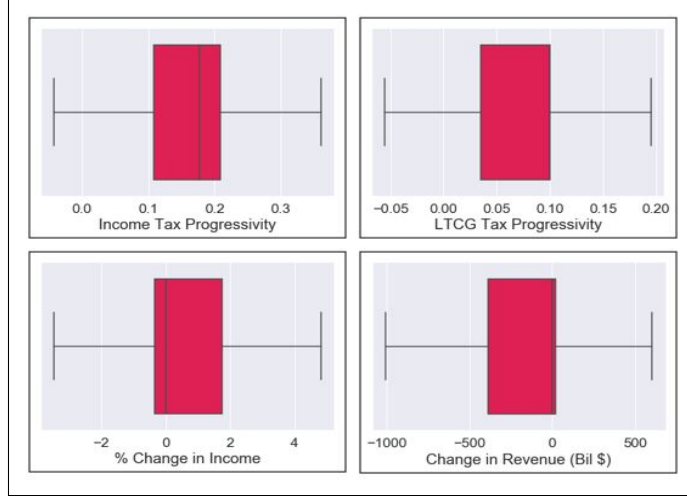


Figure 6: Distribution for computed variables

regressive tax structures [Citrin (1979), McGee et al. (2019), Ballard-Rosa et al. (2017)]. Analysis of the dataset validates these claims (at $p < 0.01$) [see Figure 7]. After further analysis, the most interesting trend is found between the subgroups of age [see Figure 8]. Age groups are labeled as young ($age \leq 35$), middle aged ($age > 35$ and $age \leq 55$), and old ($age \geq 55$). For all four computed variables, ‘young’ Republicans prefer more regressive and lower taxes as compared to the other parties. Whereas other age groups have overlapping distributions for several variables. For example, old Republicans and old Democrats have the same preference for all four variables; and middle-age Republicans prefer more regressive taxes than middle-age Democrats but both of them have the same distribution for change in revenue and change in their after-tax income.

Effect of gender: [See Figure 9] On an average, women tend to prefer higher taxes and lower after-tax income than men, but there is no effect of gender on progressivity. But when between men and women who are primary earners, there is no effect of gender on either change in revenue or change in after-tax income, and both groups favor more income and less tax. But among those who are not primary earners, women tend to prefer higher taxes and lower after-tax income.

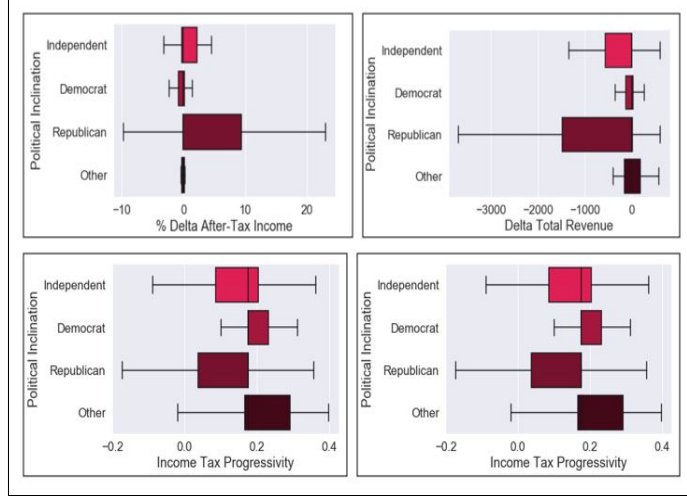


Figure 7: Effect of political affiliation

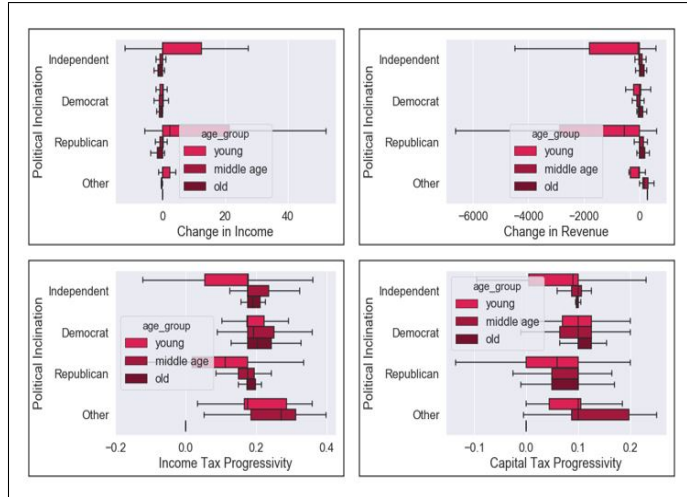


Figure 8: Effect of political affiliation and age

Finally, decision tree and random forest models are fit to predict the tax behaviour from demographic variables. [Pedregosa et al. \(2011\)](#) python package is used to implement the models. A randomized search is performed to find the optimal parameters to tune the models. Figure 10 shows the graphical representation of the optimal tree to predict LTCG progressivity from demographic variables. Table 4 shows the cross-validated Mean Squared Error (MSE) for the prediction of calculated variables from demographic variables using the two models. As expected, Random Forest models outperform Decision Tree models.

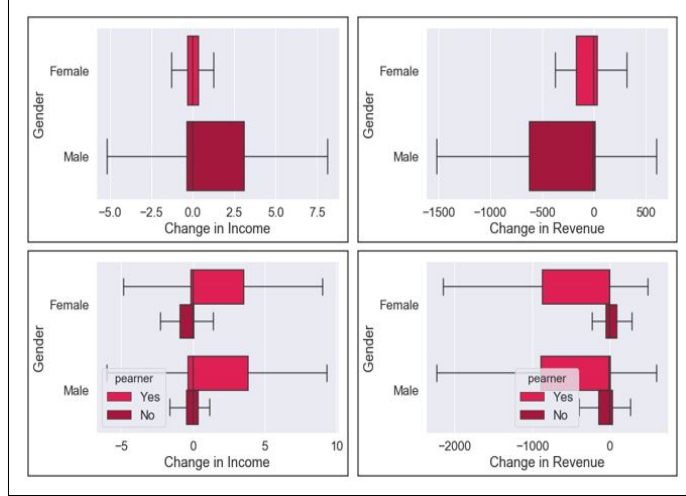


Figure 9: Effect of gender and primary earner status

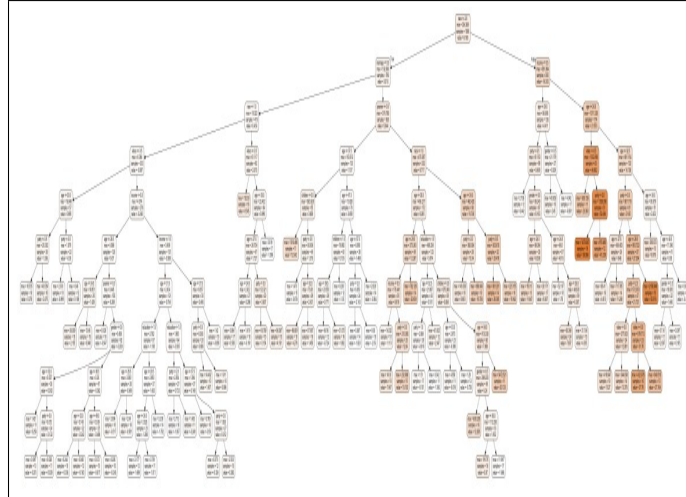


Figure 10: Optimal decision tree for fitting LTCG progressivity

5 Conclusion

The effect of the tax policy choices of respondents are calculated through the open source tax calculator simulations and through progressivity indices. While majority prefer progressive tax structures, there are significant number of people reporting both pro-tax and anti-tax preferences. While respondents seeking a tax cut, prefer a steep cut in taxes but the respondents support tax raise only want a comparatively smaller raise.

On conducting further analysis of the data several interesting trends among differ-

CV MSE	Decision Tree	Random Forest
Delta Revenue (Bil \$)	0.00155	0.00135
Delta Income (%)	286.4	256.5
PI Progressivity	0.0123	0.0119
LTCG Progressivity	0.0081	0.0078

Table 4: Cross-validated Mean Squared Error for the two models

ent sub-groups are revealed. First, the already well-established result that republicans prefer lower taxes and more regressive tax structures is verified. But upon further analysis, it is found that young republicans are the group that majorly contributing to the anti-tax preference among the party affiliates. Where as, old Republicans and old Democrats exhibit the same pro-tax behaviour. On similar lines, women show more pro-tax behaviour than men. But upon further analysis, it is found that women who are primary earners are as anti-tax as men, whereas the non primary earner women are the subgroup that exhibit pro-tax behaviour. Finally if we try to predict computed variables from demographic variable, a random forest model outperforms a decision tree model.

Going forward, a more extensive survey can yeild sufficient sample size in under-represented subgroups (e.g. old and high income republicans), so that more trends can be uncovered. Further, the new data can be used to test the predictive powers of the random forest model that was fit to the current dataset. Other statistical models (e.g. clustering) can be implimented to see if they provide a higher predictive power.

References

- Ballard-Rosa, Cameron, Lucy Martin, and Kenneth Scheve**, “The structure of American income tax policy preferences,” *The Journal of Politics*, 2017, 79 (1), 1–16.
- Beck, Paul Allen, Hal G Rainey, Keith Nicholls, and Carol Traut**, “Citizen views of taxes and services: A tale of three cities,” *Social Science Quarterly*, 1987, 68 (2), 223.
- Citrin, Jack**, “Do people want something for nothing: Public opinion on taxes and government spending,” *National Tax Journal*, 1979, 32 (2), 113–129.
- Cole, Alan**, “Details and Analysis of Donald Trump’s Tax Plan,” *Fiscal Fact No*, 2015, 482.
- de Rugy, Veronique and Jack Salmon**, “The Cost of a 70 Percent Marginal Tax Rate,” 2019.
- Gaines, Brian J**, “Flat Is Fair: American Public Opinion on Taxes and the Myth of Egalitarianism,” *The Independent Review*, 2017, 22 (1), 93–104.
- Gastwirth, Joseph L**, “A general definition of the Lorenz curve,” *Econometrica: Journal of the Econometric Society*, 1971, pp. 1037–1039.
- , “The estimation of the Lorenz curve and Gini index,” *The review of economics and statistics*, 1972, pp. 306–316.
- Horton, John J, David G Rand, and Richard J Zeckhauser**, “The online laboratory: Conducting experiments in a real labor market,” *Experimental economics*, 2011, 14 (3), 399–425.
- Jones, Eric, Travis Oliphant, Pearu Peterson et al.**, “SciPy: Open source scientific tools for Python,” 2001–. [Online; accessed <today>].

- Liaw, Andy, Matthew Wiener et al.**, “Classification and regression by random-forest,” *R news*, 2002, 2 (3), 18–22.
- McGee, Robert W, Serkan Benk, and Bahadır Yüzbaşı**, “Should Governments Tax the Rich and Subsidize the Poor? A Comparative Study of Muslim and Christian Respondents,” *Religions*, 2019, 10 (2), 72.
- Murray, Patrick**, “National: Trump’s Tax Time Troubles,” 2019.
- Nunns, Jim, Len Burman, Jeffrey Rohaly, and Joseph Rosenberg**, “An analysis of Donald Trump’s revised tax plan,” *Washington, DC: Tax Policy Center*. [http://www. taxpolicycenter. org/publications/analysisdonald-trumps-revised-tax-plan](http://www.taxpolicycenter.org/publications/analysisdonald-trumps-revised-tax-plan), 2016.
- Owen, DB**, “The power of Student’s t-test,” *Journal of the American Statistical Association*, 1965, 60 (309), 320–333.
- Paolacci, Gabriele, Jesse Chandler, and Panagiotis G Ipeirotis**, “Running experiments on amazon mechanical turk,” *Judgment and Decision making*, 2010, 5 (5), 411–419.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay**, “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, 2011, 12, 2825–2830.
- Pew Research Center**, “Tax system seem as unfair, in need of overhaul,” 2011.
- , “U.S. income tax is progressive, but enough to be ‘fair’?,” 2017. Accessed: 2019-11-06.
- , “Public Has Mixed Expectations for New Tax Law,” 2018.
- , “Growing Partisan Divide Over Fairness of the Nation’s Tax System,” 2019.

- Quinlan, J. Ross**, “Induction of decision trees,” *Machine learning*, 1986, *1* (1), 81–106.
- Roberts, Michael L, Peggy A Hite, and Cassie F Bradley**, “Understanding attitudes toward progressive taxation,” *Public Opinion Quarterly*, 1994, *58* (2), 165–190.
- Saez, Emmanuel and Gabriel Zucman**, “Alexandria Ocasio-Cortez’s Tax Hike Idea Is Not about Soaking the Rich,” *New York Times*, 2019.
- Suits, Daniel B**, “Measurement of tax progressivity,” *The American Economic Review*, 1977, *67* (4), 747–752.
- Tax-Calculator 2.2.0**, “Tax-Calculator 2.2.0, author’s calculations,” <https://PSLmodels.github.io/Tax-Calculator> 2019. Accessed: 2019-11-06.
- World Bank**, “Comprehensive Taxpayer Attitude Survey (CTAS),” 2017.
- , “Of Governance and Revenue: Participatory Institutions and Tax Compliance in Brazil,” 2019.

APPENDIX

A-1 Survey Design Variables

Survey variables	Bracket	Current Value	New Values Range
Maximum taxable earnings for Social Security	-	\$128,400	Any number
LTCG Tax 1	\$0 to \$37,950	0%	0% to 100%
LTCG Tax 2	\$37,951 to \$418,400	15%	0% to 100%
LTCG Tax 3	\$418,401 +	20%	0% to 100%
PI Tax 1	\$0 to \$9,700	10%	0% to 100%
PI Tax 2	\$9,701 to \$39,475	12%	0% to 100%
PI Tax 3	\$39,476 to \$84,200	22%	0% to 100%
PI Tax 4	\$84,201 to \$160,725	24%	0% to 100%
PI Tax 5	\$160,726 to \$204,100	32%	0% to 100%
PI Tax 6	\$204,101 to \$510,300	35%	0% to 100%
PI Tax 7	\$510,300 +	37%	0% to 100%

Table 5: Tax survey variables design

Survey variables	Input Values
Gender	Male or Female or Other
Age	Any number
Political Affiliation	Democrat or Independent or Republican or Other
Education	Less than high school High school Some college 2 year degree 4 year degree Professional degree Doctorate
Race	American Indian or Alaska Native Asian Black or African American Hispanic or Latino Native Hawaiian or Pacific Islander White Other Don't know or not sure
Annual Household Income, 2018	Less than \$20k \$20k - \$40k \$40k - \$75k \$75k - \$100k \$100k - \$200k \$200k - \$500k More than \$500k
Primary earner of the household	Yes or no
Marital status	Married or not married
No. of children	0, 1, 2, and 3 or more

Table 6: Demographic survey variables design