Fairness and Tax Preferences: A Conjoint Experiment *Pre-Analysis Plan*

Mariana Alvarado

October 16, 2017

1 Background

In light of rising levels of inequality, the last few years have witnessed growing interest and research on issues of taxation and redistribution. At the policy level, the question of tax reform has garnered increased attention and led to deep partisan disagreements, mainly regarding how much taxes the rich should pay[1](Ballard-Rosa et al., 2017). Beyond these political battles however, policy-makers face the challenge of designing a tax system that can be supported by voters on both sides of the political spectrum.

In terms of research, studies examining Americans' preferences regarding taxation have consistently found that people not only care about maximizing their own income, but also about fairness (Alesina et al., 2012). And research regarding fairness in taxation has focused mostly on how tax preferences vary on the basis of perceived deservingness. In that sense, both survey (Alesina and Angeletos, 2005; Fong, 2001) and experimental (Chow and Galak, 2012; Durante et al., 2014) evidence have consistently found that people prefer income or wealth earned through (or perceived as having been earned through) luck to be taxed at a higher rate than income or wealth that result from (or are perceived as resulting from) effort. Other approaches focus on establishing the relationship between people's individual attitudes towards inequality and their tax preferences (Lü and Scheve, 2016; McCall et al., 2017).

[1]See for example The Wall Street Journal, "The GOP's 'Tax the Rich' Temptation, October 8th 2017, https://www.wsj.com/articles/the-gops-tax-the-rich-temptation-1507487704 (accessed October 13th 2017), and The New York Times, "I.M.F. Cautions Against Tax Cuts for the Wealthy as Republicans Consider Them", October 11th 2017, https://www.nytimes.com/2017/10/11/us/politics/imf-tax-cuts.html (accessed October 13th 2017).

This research agenda suffers from two important shortcomings. In the first place, there is a conspicuous gap between the political debate -which has focused on the relationship between income level and taxes- and academic research -which has focused on deservingness, or the relationship between income source and taxes-. Secondly, academic research has tended to confound the questions of who should pay more taxes and why they should pay more taxes. However, in order to find policy solutions that respond to public opinion, a more comprehensive assessment of peoples' fairness considerations that overcomes these two shortcomings is necessary.

This project seeks to provide such an assessment by conducting a conjoint experiment that will be able to isolate the relative importance of level and source of income in determining peoples' preferences for taxation.

2 Project Description

I will conduct a conjoint experiment that will allow me to identify what it is people take into consideration when deciding how to distribute the tax burden. Conjoint survey experiments "ask respondents to choose from or rate hypothetical profiles that combine multiple attributes, enabling researchers to estimate the relative influence of each attribute value on the resulting choice or rating" (Hainmueller et al., 2014, p. 2). In this case, respondents will be presented with a pair of profiles in which income level, source of income, and proportion of income paid in sales taxes will be randomly varied, and ask to choose which one of these profiles should pay a higher tax rate. This design offers several advantages. The fact that attributes vary randomly will allow me to disentangle the effects of correlated attributes (e.g., wealth and luck). The forced choice component neutralizes attitudes about the overall level of taxation, so I can focus on the attributes that make citizens appear as more or less taxable to the respondent. Furthermore, since the resulting estimates represent effects on the same outcome (in this case, the probability that a profile will be chosen to receive the higher tax rate), they can be compared in order to assess the relative influence of different attributes. Finally, this design will also allow me to assess the existence of heterogeneity in preferences by respondent characteristics (e.g., income level, inequality aversion), or the extent to which people combine different fairness considerations when making their decisions.

As mentioned above, respondents will have to choose between a pair of profiles in which income level, source of income and proportion of income paid in sales taxes will vary randomly. Income and percentage of income paid in taxes will vary discretely from low to high, while sources of income will represent either effort, luck, social background or state privilege. See table 1 for attribute levels.

Attribute	**Attribute levels**
Annual income	\$30,000, \$80,000, \$150,000
Source of income	Started own small business, Receives annuity from a lottery prize, Owns business that was bailed out by government, Got a job through family connections
% of income paid in sales taxes	1%, 5%, 10%

Table 1: Attributes and Attribute Levels

In terms of presentation, two profiles will be presented side-by-side on the same screen, with the following prelude:

> Many observers in the United States have discussed the possibility of changing the federal income tax code to address multiple issues. The design of a new tax system raises a number of questions, including whether and why some people should pay higher rates than others. We are interested in what you think about this.
>
> We will show you profiles of random individuals. You will be shown pairs of individuals, along with several of their attributes. For each comparison we would like to know which of the two individuals you think should pay a higher tax rate. In total, we will show you five comparison pairs.
>
> Bear in mind that when we talk about tax rates we mean the percentage of their income that someone pays in taxes. People with different incomes who pay the same rate actually pay different amounts (i.e., 30% of an income of $100,000 is \$30,000, but of an income of $50,000 it is \$15,000).
>
> Please take your time when reading the attributes of each individual. People have different opinions about this issue, and there are no right or wrong answers.

This introduction will be followed by a screen similar to figure 1.

Attributes	Individual 1	Individual 2
Annual income	$150,000	$80,000
Source of income	Started own small business	Receives annuity from lottery prize
Percentage of income paid in sales taxes	1%	5%

Which of the two individuals would you personally prefer to charge a higher tax rate to?

Individual 1 | Individual 2

Figure 1: Example of choice-based conjoint survey

In order to maximize the number of observations and allow respondents to familiarize themselves with the format of the experiment, each subject will see 5 pairs of profiles. After the first pair of profiles, they will be asked to answer the following open ended question, which will provide supplementary information on respondents' fairness preferences:

> Why did you choose citizen <chosen citizen>?

After completing their 5 choice tasks, respondents will have to fill a survey asking for their socio-demographic information (age, gender, education, household income, partisanship, employment status, race, marital status, ideology and zip code of residence), which will be used to check for heterogeneous preferences. They will also be asked to answer a question regarding their general preferences for progressivity, which will be used as a validation check, and a question regarding their opinion about current levels of inequality, which will be used to determine whether they are inequity averse:

> Do you think everyone should pay the same share of their income in taxes or some people should pay a higher share than others?
>
> American households with incomes in the top 10% earn an average of $230,000 per year, and households with incomes in the bottom 50% earn an average of $25,000 per year. Should this difference be bigger, smaller, or about what it is now?

Although no randomization restrictions will be used, I will test whether responses vary for profiles that might be considered atypical[2]. In order to be able to rule out primacy and recency effects, the order in which the attributes appear in the screen will be randomized between subjects[3]. I will also implement an attention test[4] after 3 of the 5 pairs of profiles have been shown, and record the time it takes each respondent to complete the whole survey and to choose each profile. Both of these strategies will allow me to test whether choices vary on the basis of attention paid.

3 Sample

This conjoint survey experiment will be completed online using Amazon's Mechanical Turk (MTurk) by a sample of 2,000 respondents. Respondents will be required to be US residents and have an approval rate above 90%. The survey takes approximately 6 minutes, and respondents who complete it will be paid 60 cents.

The job will be published on MTurk in several batches, with the requirement that respondents who participated in a previous batch cannot participate again.

4 Models

Outcome data will come from the forced choice made by respondents regarding which profile in each pair should pay a higher tax rate. The unit of analysis is the citizen profile and outcomes will be measured using a dummy variable that will take a value of 1 if a profile is chosen and a 0 if a profile is not chosen.

All results will be illustrated graphically using plots to show the effect of each attribute value on the probability of a profile being chosen to pay the highest rate, relative to the baseline attribute value, with 95% confidence intervals.

For the analyses below I will exclude respondents who completed the survey in less than half of the median time. I will also check if respondents compared the exact same profile in any of their tasks and if so, exclude those observations.

[2]Atypical profiles would be: "Receives annuity from a lottery prize" and $30,000; "Owns business that was bailed out by government" and $30,000.

[3]Not within subjects, to reduce complexity.

[4]Following Bechtel and Scheve (2013), the attention test will involve asking respondents the following question: "We are interested in learning about your preferences on a variety of topics, including colors. To demonstrate that you've read this much, just go ahead and select both red and green among the alternatives below, no matter what your favorite color is. Yes, ignore the question below and select both of those options. What is your favorite color?"

4.1 Main Analysis

My benchmark regression model will estimate the Average Marginal Component Effects (AMCEs) for each attribute value, pooling across all respondents. AMCEs measure the average change in the probability that a profile will be chosen to have the higher tax rate when it includes a given attribute value compared to when it includes the baseline attribute value.

AMCEs can be estimated via a single[5] OLS regression of the outcome variable on dummies for each level of each attribute (excluding a reference category[6]), (Hainmueller et al., 2014), as shown in equation 1. Subscripts i identify respondents, j identify the profile in each pair ($j = 1, 2$), and k identify the choice task or profile pair ($k = 1, 2, 3, 4, 5$).

$$\begin{aligned} chosen_{ijk} &= \beta_0 + \beta_1 I2_{ijk} + \beta_2 I3_{ijk} + \beta_3 S2_{ijk} + \beta_4 S3_{ijk} + \beta_5 S4_{ijk} \\ &+ \beta_6 T2_{ijk} + \beta_7 T3_{ijk} + \varepsilon_{ijk} \end{aligned} \tag{1}$$

The equation above will be estimated through an OLS regression with standard errors clustered by respondent to account for within-respondent correlation.

Given that my design did not include any randomization restrictions, calculating the AMCEs is very straightforward: the AMCE for each attribute level is equal to the estimated coefficients on their dummies (Hainmueller et al., 2014).

4.2 Heterogeneous effects

I will also estimate conditional AMCEs to see if the causal effect of an attribute interacts with respondents' background characteristics. I will do so by interacting attributes and respondent characteristics to test whether AMCEs differ between groups; for graphical purposes I will also conduct subgroup analysis and present plots with results for different groups side by side.

1. By inequality aversion 1: compare estimates between those who think inequality is too high and those who think it is too low or ok.

2. By inequality aversion 2: compare estimates by level of inequality (gini coefficient) in zip code. Create 3 groups based on level of inequality in zip code

[5] A single regression can be used to estimate all effects because the regressors are in expectation orthogonal to one another.

[6] Excluded baseline attribute values: \$30,000 annual income, 10% paid in sales taxes, Started own small business (effort).

(low=deciles 1-3, medium=deciles 4-7, high=deciles 8-10)[7] and compare estimated AMCEs across these three groups. Do the same at state instead of zip code level[8].

3. By income level 1: compare estimates between respondents with low (below median in sample) vs high (above median in sample) income.

4. By income level 2: compare estimates between respondents with low (under $30,000), medium (between $30,000 and $80,000) and high (over $80,000) levels of income.

5. By income level 3: compare estimates between respondents with incomes within $15,000 margins around levels of income presented: $15,000-$45,000 vs $65,000-$95,000 vs $135,000-$165,000.

6. By race: white vs non-white.

7. By sex: males vs females.

8. By marital status: single vs ever been married (divorced, separated, widowed, married).

9. By partisanship: democrats vs republicans.

10. By ideology: conservatives vs liberals.

11. By vote choice 2016: Trump vs Clinton.

12. By level of education: low education (up to high-school degree) vs high education (above high-school degree). Also low (up to high school degree), middle (above high school and below post-graduate), and high (MA and over).

13. By citizenship: born here vs not born here.

14. By age: high (above median) vs low (below median). Also by age groups: 29 and under, 30-39, 40-49, 50-59, 60 and over.

15. By employment status: by each response option (full-time employee, part-time employee, self-employed or small business owner, unemployed and looking for work, student, not in the labor force). Also employed vs unemployed vs not in workforce (including students).

[7] Zip-code level gini coefficients from the 2011-2015 American Community Survey 5 year estimates.

[8] State-level gini coefficients from the 2016 American Community Survey 1-year estimates.

16. By preferences for progressivity: compare estimates between those who think everyone should pay the same tax rate and those who think people should pay different rates. Describe average respondent who thinks people should pay same rate and average respondent who thinks people should pay different rates. Check if demographic attributes predict responses (regress response on demographics).

4.3 Interactions between attributes

I will estimate average component interaction effects (ACIEs) to see if an attribute interacts with another attribute to influence responses. Again, I will interact attribute values to test whether AMCEs differ; for graphical purposes I will conduct subgroup analysis and present plots with results by different levels of one of the attributes.

- Check for interactions between share of income paid in sales taxes and income level.
- Check for interactions between level of income and source of income.
- Check for interactions between source of income and share of income paid in sales taxes.

4.4 Diagnostic checks

These are intended to check that the conjoint analysis assumptions hold and to probe potential concerns about external validity (row-order effects and atypical profiles). These are all based on Hainmueller et al. (2014).

- Carryover effects: estimate AMCEs separately for each of the 5 rounds of choices. Assumption 1 (stability and no carry-over effects assumption) implies AMCEs should be similar across tasks.
- Profile order effects: estimate AMCEs separately for each profile (1 and 2). Assumption 2 (no profile order effects) implies AMCEs should be similar regardless of whether the attribute occurs in the first or second profile in a given task.
- Randomization: conduct multivariate balance checks by regressing respondent characteristics on indicator variables for all profile attributes.

- Row-order effects: estimate row-specific AMCEs and test whether the estimates are significantly different from each other. This is to examine if the AMCE of an attribute depends on the order in which it appears in the conjoint table.

- Atypical profiles: classify profiles by their typicality (as described in footnote 2) and then estimate the AMCEs for groups of respondents exposed to different number of atypical profiles (e.g. low 0-3, medium 4-5 and high 6-9). Note that atypical profiles are included because it is precisely by breaking the correlations that exist in reality that we can separate the isolated effect of each attribute (Hainmueller et al., 2014).

4.5 Robustness tests

- Repeat analyses separately for those who passed and didn't pass the attention screener.

- Check if screening inattentives leads to selection bias.

- Assess results separately by time of survey completion (below and above median).

- Repeat analyses excluding respondents who participated in formative study.

- Include false discovery rate (FDR) p-value adjustment for multiple testing in analyses of heterogeneous effects using different operationalizations of respondents' income levels, level of education, age, and employment status.

- Repeat analyses using weights to adjust the sample such that it matches the demographic and geographic margins of the adult population (using the March 2016 Supplement to the Current Population Survey) on age, gender, education, race, household income, census region and partisanship. Weights will be created using entropy balancing.

5 Missing data

Observations with missing outcome data (i.e., respondents did not choose any of the two profiles) will be dropped. Observations with missing data on covariates will be kept in the main analyses but dropped in the estimation of heterogeneous effects.

References

Alberto Alesina and George-Marios Angeletos. Fairness and redistribution. *The American Economic Review*, 95(4):960–980, 2005.

Alberto Alesina, Guido Cozzi, and Noemi Mantovan. The evolution of ideology, fairness and redistribution. *The Economic Journal*, 122(565):1244–1261, 2012.

Cameron Ballard-Rosa, Lucy Martin, and Kenneth Scheve. The structure of american income tax policy preferences. *The Journal of Politics*, 79(1):000–000, 2017.

Michael M Bechtel and Kenneth F Scheve. Mass support for global climate agreements depends on institutional design. *Proceedings of the National Academy of Sciences*, 110(34):13763–13768, 2013.

Rosalind M Chow and Jeff Galak. The effect of inequality frames on support for redistributive tax policies. *Psychological science*, 23(12):1467–1469, 2012.

Ruben Durante, Louis Putterman, and Joël Weele. Preferences for redistribution and perception of fairness: An experimental study. *Journal of the European Economic Association*, 12(4):1059–1086, 2014.

Christina Fong. Social preferences, self-interest, and the demand for redistribution. *Journal of Public economics*, 82(2):225–246, 2001.

Jens Hainmueller, Daniel J Hopkins, and Teppei Yamamoto. Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, 22(1):1–30, 2014.

Xiaobo Lü and Kenneth Scheve. Self-centered inequity aversion and the mass politics of taxation. *Comparative Political Studies*, 49(14):1965–1997, 2016.

Leslie McCall, Derek Burk, Marie Laperrière, and Jennifer A Richeson. Exposure to rising inequality shapes americans opportunity beliefs and policy support. *Proceedings of the National Academy of Sciences*, 114(36):9593–9598, 2017.