Tax evasion: Experimental lab evidence from India

PRE-ANALYSIS PLAN

NOVEMBER 2021

PRIMARY INVESTIGATORS

Dr Sanjit Dhami, University of Leicester

Dr Narges Hajimoladarvish, Ashoka University

Dr Pavan Mamidi, Ashoka University

CONTRIBUTORS

Aayush Agarwal, Ashoka University

Bijoyetri Samaddar, Ashoka University

This project is officially documented here, on the AEA RCT Registry.

1. Introduction and Background

1. This experiment measures loss and risk aversion in the context of entitled and unentitled income in order to understand tax evasion with relation to mental accounting.
2. Participants will be asked to make decisions regarding paying taxes while taking into account various factors like tax, audit, and fine rates, and choosing between a lottery and a sure amount, wherein the lottery may or may not increase their payout.
3. There are four treatment arms:
    a. Unprimed group: Group that is not shown data on tax evasion in India and earn their income
    b. Primed group: Group that is shown data on tax evasion in India and earn their income
    c. Unprimed group: Group that is not shown data on tax evasion in India and receive a random income
    d. Primed group: Group that is shown data on tax evasion in India and receive a random income
4. Depending on which treatment group they have been assigned to, participants are exposed to the prime and then asked to make decisions.
5. Participants make decisions about their labour (earned)/ non-labour (random) income based on the treatment they are assigned to.
6. Finally, participants are asked to fill out a post-experiment questionnaire where we record their demographics.
7. This experiment will be implemented on zoom.

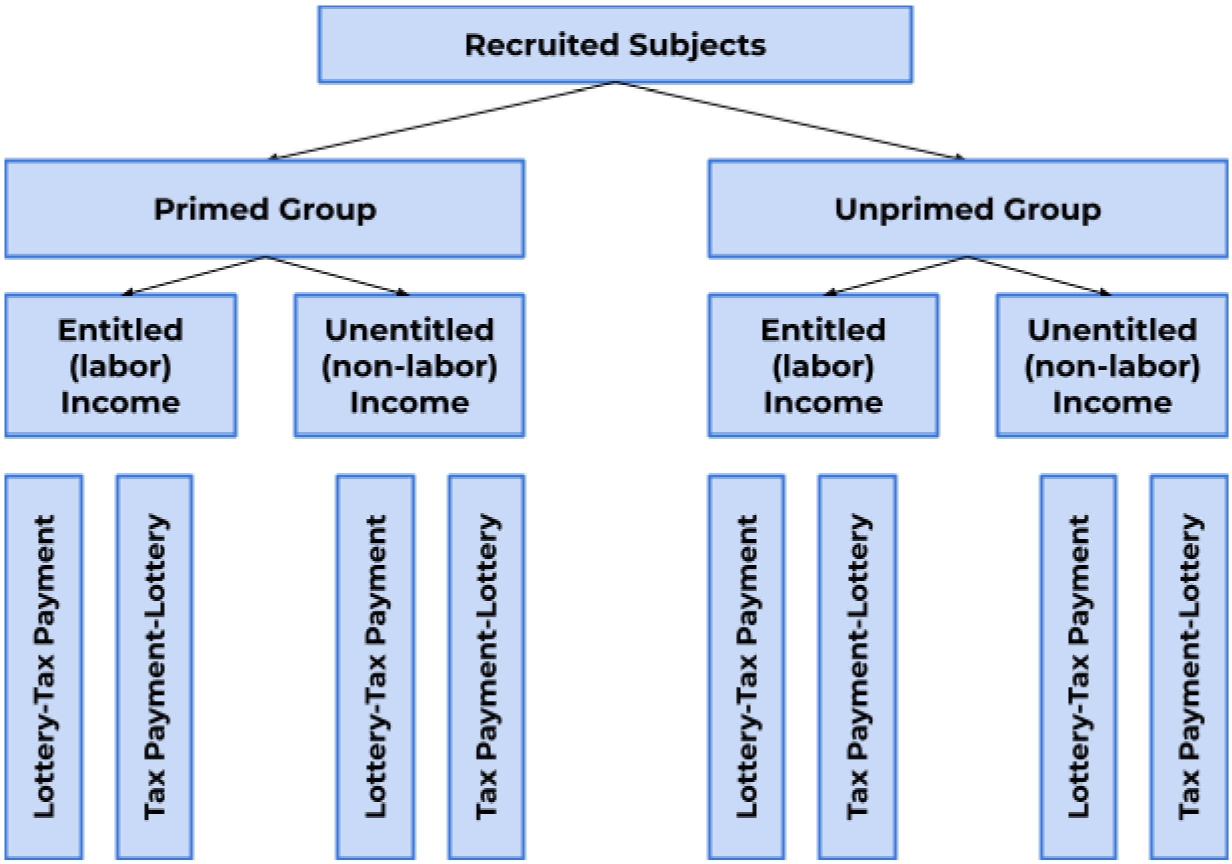
This document outlines the analysis plan for the experiment.

2. Methods

2.1. Experimental design overview

We are interested to investigate if tax evasion is related to source of income (Labour and non-labour income) and behavioural parameters such as loss aversion as predicted by our model.

Our experiment design has multiple layers of randomization that take into account the various effects we are interested in exploring. We first randomly assign subjects into the primed group and the unprimed group, wherein the primed group will be exposed to data on tax evasion in India and the unprimed group will not be exposed to that data. We are also interested in distinguishing between the effects of the type of income source, hence subjects are further assigned between the entitled/earned income arm and the unentitled/random income arm. Finally, we take into account task order randomization between the lottery (loss aversion) task and the tax payment task to control for any possible order effects.



2.2. Sample Identification

We want tax-paying subjects, preferably those who do not have tax deducted at source but rather have to file their own tax returns. We want a sample that is fairly representative of real-world taxpayers and their different kinds of occupations. This includes occupations such as:

1. Doctors
2. Hedge fund managers
3. Lawyers
4. Business owners
5. Shopkeepers
6. Chartered accountants
7. Company secretaries
8. Any other professionals who charge fees
9. Self-employed individuals

Sample selection criteria include:

- 18+ years of age (preferably 24+)
- Access to internet

- Employed either part or full time
- Have filed income tax returns before (TDS not counted)
- Comfortable with either English or Hindi

* Due to problems with finding such a sample, we had to run the experiment with students.

2.3. Data Collection

We will be conducting on-field data collection by contracting with an enumerator agency. Enumerators will conduct random sampling in specific parts of Delhi, including marketplaces, courts, educational institutes, and other relevant areas. Enumerators will hand over their tablets to the respondents, hence the experiment will be self-administered.

* Due to problems with finding such a sample, we had to run the experiment with students over zoom.

3. Empirical Analysis

3.1. Variables

For each individual we elicit loss aversion and risk aversion. We also ask individuals to declare their (Labour or non-labour) income for tax purposes with varying tax, audit, and fine rates.

Outcome Variable	Description	Outcome Measures	Outcome Measure Type
Income: y	This is determined either randomly in non-labour income treatment arm or based on subject's performance in effort tasks in labour income treatment arm.	Minimum income = 25 Maximum income = 125	Categorical variable [25, 50, 75, 100, 125]
x	This is an amount required to calculate loss aversion parameter. This is explained in section 4.1 of the original paper.	**x** is calculated through 6 binary choices between lotteries and sure amounts. These 6 choices are part of the bisection procedure to determine the value of x such that $(y - z, 0.5; y + x; 0.5) \sim y$	Continuous variable [0,5z]

C	This is an amount required to calculate risk aversion and parameter of utility function.	$(y,\ 0.5;\ 2y\ ;\ 0.5) \sim C$	Continuous variable [y, 2y]
dl1	Subject's declared income when tax rate is **5%**, fine rate is 2 and probability of audit is 3%.	This is self- reported subject's declared income for tax purposes	Continuous variable [0,y]
dl2	Subject's declared income when tax rate is **30%**, fine rate is 2 and probability of audit is 3%.	This is self- reported subject's declared income for tax purposes	Continuous variable [0,y]
dl3	Subject's declared income when tax rate is **60%**, fine rate is 2 and probability of audit is 3%.	This is self- reported subject's declared income for tax purposes	Continuous variable [0,y]
fl1	Subject's declared income when tax rate is 30%, fine rate is **0** and probability of audit is 3%.	This is self- reported subject's declared income for tax purposes	Continuous variable [0,y]
fl2	Subject's declared income when tax rate is 30%, fine rate is **1** and probability of audit is 3%.	This is self- reported subject's declared income for tax purposes	Continuous variable [0,y]
fl3	Subject's declared income when tax rate is 30%, fine rate is **2** and probability of audit is 3%.	This is self- reported subject's declared income for tax purposes	Continuous variable [0,y]
adl1	Subject's declared income when tax rate is 30%, fine rate is 2 and **probability of audit is decreasing in declared**	This is self- reported subject's declared income for tax purposes	Continuous variable [0,y]

	income.		

Table 1: Primary Outcome Measures

After eliciting **x** and **c** we calculate loss aversion and risk aversion as follows. Using prospect theory and assuming reference point equal to income, the indifference between y and $(y - z,\ 0.5;\ y + x;\ 0.5)$ with power utility gives:

$$w(0.5)\,x^{\gamma} - w(0.5)\lambda z^{\gamma} = 0$$

Thus,

$$\lambda = \frac{w(0.5)}{w(0.5)}\left(\frac{x}{z}\right)^{\gamma} = \left(\frac{x}{z}\right)^{\gamma}$$

The only unknown in the above formula is γ, which is calculated using the indifference between C and $(y,\ 0.5;\ 2y;\ 0.5)$. Again assuming reference point equal to income, we have

$$v(C - y) = [1 - w(0.5)]v(y - y) + w(0.5)v(2y - y)$$

$$v(C - y) = w(0.5)v(y)$$

$$(C - y)^{\gamma} = w(0.5)y^{\gamma}$$

Using the approximation, $w(0.5)\approx 0.5$, we get the following approximation for γ that we use in our estimates.

$$\gamma \approx \frac{log(0.5)}{log(\frac{c-y}{y})}$$

Given the noisy nature of these estimates, we shall use binary measures of loss aversion and risk aversion. A subject is classified as loss averse if λ > 1, otherwise she is loss tolerant (λ < 1). Similarly, if $1.5y - C > 0$ subject is classified as risk averse and otherwise with $1.5y - C < 0$, she is classified as risk loving.

dl1-dl3 and **fl1-fl3** will be used to test the comparative statistics results when we vary the tax rate and the penalty rate. We use **adl1** for treatment effects analysis as it has more realistic values of the policy parameters.

The following variables will be used as covariates in regression analysis.

Type	**Description**	**Measure creation/ Final variable used**
Covariates: Demographics	Please enter your age:	Numeric entry
	Please select your gender:	1. Female

		2. Male 3. Transgender 4. Other (please specify): _____
	Marital Status	List entry
	Please select your degree of study:	1. Economics 2. Psychology 3. BBA/MBA/Statistics 4. Liturture 5. Sociology 6. Physics/Chemistry/Biology 7. Engineering 8. Any other 9. Not applicable
	Employment Status	1. 1. Employed full-time 2. 2. Employed part-time 3. 3. Self-employed 4. 4. Unemployed 5. 5. Student 6. 6. Others
	Are you the only earning member in your household?	1. Yes 0. No
	Besides yourself, how many earning members are in your household?	<<integer>>
	Which religion do you follow?	1. Hinduism 2. Islam 3. Christianity 4. Jainism 5. Sikhism 6. Other (please specify): _____
	money spent on household expenses in last 7 days	integer

Table 2: Secondary Outcome Measures

3.2. Model Specifications

We run regressions using our primary outcome measures and calculated loss aversion and risk aversion, with and without our control variables. We will either use an OLS model or an Ordered Probit Model, depending on the distribution of variables and number of corner solutions. The dependent variable is income minus declared income for taxes ($y - adl1$). If many subjects choose to declare no income or full income, we will use Ordered Probit Model. Otherwise we shall use an OLS model. We also run regressions with continuous measures for loss aversion and risk aversion.

We are interested in (i) the determinants of tax evasion, particularly loss aversion, which is the key transmission mechanism in all non-expected utility theories of evasion, but has never been directly confirmed, and (ii) the treatment effects. Let $Z = y - adl1$ denotes evaded income.

Following an ordered probit analysis requires first specifying a desired, or latent, level of evasion, Z^*, for a taxpayer. This is given by

$$Z^* = \alpha_0 + \alpha_1 d_\lambda + \alpha_2 d_R + \alpha_3 d_T + \alpha_4 d_\lambda d_T + \alpha_5 d_\lambda d_R + \alpha_6 d_R d_T + \alpha_7 d_\lambda d_R d_T + \delta X_D + u$$

Using more compact notation we have

$$Z^* = \beta X + u$$

In the case of OLS we estimate the above model, where $Z^* = Z$. In this model u is normally distributed with mean zero and regressors are as follows.

- d_λ or "Loss aversion" is a dummy variable that takes the value 1 if the subject is loss averse ($\lambda > 1$) and 0 if the subject is loss tolerant ($\lambda < 1$).
- d_R or "Risk aversion" is a dummy variable that captures attitudes to risk. It takes the value 1 for risk averse individuals and 0 for risk loving individuals.
- d_T or "Labor" is a treatment dummy that equals 1 for Treatment 1 (labor income) and 0 for Treatment 2 (non-labor income).
- d_p is a treatment dummy that equals 1 for primed groups and 0 for unprimed groups.
- We also have the following interaction terms between various dummy variables: $d_\lambda d_T$, $d_\lambda d_R$, $d_R d_T$, $d_\lambda d_R d_T$.
- The vector X_D contains a number of subject-specific control variables, mostly of a demographic nature, that we now explain:
- Income' is the subject's income from the experiment. It has 5 categories, one each for the income levels 25, 50, 75, 100, 125.
- Time' indicates the length of time taken for the completion of the experiment.
- Age' gives the self-reported age of subjects.
- Gender' is a dummy for gender and takes values 1 for male and 0 for female.

- Marital' is a dummy for marital status and takes values 0 for single and 1 otherwise.

Next, we define 4 binary variables, Z_k, $k = 0, 1, 2, 3$ for any taxpayer with income y.

$Z_0 = \{1\ if\ Z = 0\ \ 0\ otherwise$ $\qquad Z_1 = \{1\ if\ 0 < Z < \frac{1}{2}y\ \ 0\ otherwise$

$Z_2 = \{1\ if\ \frac{1}{2}y \leq Z < y\ \ 0\ otherwise$ $\qquad Z_3 = \{1\ if\ Z = y\ \ 0\ otherwise$

These binary variables are related to the latent variable Z^* as follows.

$$Z_0 = 1\ if\ Z^* < c_1;\ Z_1 = 1\ if\ c_1 < Z^* < c_2;\ Z_2 = 1\ if\ c_2 < Z^* < c_3;\ Z_3 = 1\ if\ c_3 < Z^*$$

Using our assumptions on the error term u and the above expressions, denoting probability by P, and the cumulative distribution function of the normal distribution by ϕ, we can now find the following expressions that enable us to use ordered probit model and compute the marginal probability effects.

$$P\left(Z_0 = 1\right) = P\left(Z^* < c_1\right) = P\left(u < c_1 - \beta X\right) = \phi(c_1 - \beta X)$$

$$P\left(Z_1 = 1\right) = P\left(c_1 < Z^* < c_2\right) = \phi\left(c_2 - \beta X\right) - \phi(c_1 - \beta X)$$

$$P\left(Z_2 = 1\right) = P\left(c_2 < Z^* < c_3\right) = \phi\left(c_3 - \beta X\right) - \phi(c_2 - \beta X)$$

$$P\left(Z_3 = 1\right) = P\left(c_3 < Z^* < c_4\right) = 1 - \phi(c_3 - \beta X)$$