ENV ECON 118 / IAS 118 - Introductory Applied Econometrics
Assignment 3
Due Tuesday October 13 at the beginning of class

Exercise 1: Efficient Targeting of Public Programs

Background

The data for this exercise come from the paper "Private Information and the Allocation of Land Use Subsidies in Malawi" by Kelsey Jack, published in July 2013 in *The American Economic Journal: Applied Economics*. Health, environment, and social safety net programs are often designed to target resources towards individuals who will derive the greatest benefits from the programs in question. How are these people identified? Traditionally, implementing partners have used techniques that induce self-selection into the program. This means that systems are set up so that only those who truly need the resources (and will make proper use of them) actually claim the support. Inducing self-selection is meant to increase efficiency because each potential recipient has private information about their own benefits and costs under these programs, which public agencies cannot observe. In practice how are these set up? One example of a mechanism that induces self selection is an auction: individuals are asked to place a bid with their willingness to pay for a subsidized good in an envelope. The NGO/implementing partner will then select all the bids above a certain price point.

Despite the importance of these programs, very few field studies have investigated how targeting public funds through self-selection affects program outcomes. Are they better than just simply randomly allocating the public funds to members within the community via a lottery system for example? This is what Kelsey Jack sets out to test, and what we investigate in this problem set.

To this end, Jack implements a randomized control trial (RCT) in central Malawi. Together with the World Agroforestry Centre (ICRAF), they randomly assigned households to one of two treatments arms for a program that subsidizes afforestation on private land. The program gives landholder recipients 50 seedlings, to be planted on half an acre of land, and pay them for each surviving tree at regular intervals over a period of three years. This type of land-use subsidy is commonly used by governments because it has both private benefits (increases soil fertility, income from timber), and public benefits (carbon sequestration, reduced erosion, and biodiversity). The two treatment arms are:

1. **Auction**: Landowners were asked to bid the lowest price for surviving trees they would be willing to accept to participate in the scheme. The program selected as participants those that bid the lowest prices.

2. **Lottery**: Participants were randomly chosen among those that were invited.

Chosen participants in both treatments received seedlings, and tree survival outcomes were measured two years later. During the same visit, landowners also received a score (1 to 5) assessing the quality of maintenance on their fields.

Data

The data include observations on all 433 individuals that participated in the process. This includes those that were allocated to the lottery group, and those that were allocated to the auction group. Note, that not everyone in these two groups received seedlings. In the bidding treatment this is because some people's bids were too high. In the lottery treatment this is because some people simply weren't selected. The dataset posted on the course website contains the following variables:

- respondentid: Unique identifier for the landowner
- auction: =1 if in auction treatment
- educ: Education of respondent
- hhsize: Household size
- acres_total: Landholding (acres)
- income_ganyu: Household has casual labor income (0/1)
- trees: Prior tree planting experience (0/1)
- assets_sum: Asset index
- icraf: Prior contact with ICRAF (0/1)
- outsiders: Index of mistrusts of outsiders (1-3)
- trynew: index of willingness to try new technology (1-3)
- alive: Number of surviving trees
- score: Field quality score in visits

1. Briefly describe your data set. Since these data investigate the use of two different methods, you a) want to look at the number of individuals assigned to each group (and the number in the sample overall), and b) compare the means of at least 4 characteristics of these individuals across the two groups: average education, casual labor income, assets, total acres. By compare I mean discuss whether they are similar across groups, and why you might expect such a result.

2. We will now compare these two targeting methods by looking at tree survival outcomes across the two groups. Draw a histogram of number of surviving trees in each group. Comment on differences (be precise - and explain why the differences intuitively make sense).

You should not use STATA for questions 3-6, except to calculate the sample mean, variance, and # obs. Show all work

3. Compute an estimate for the mean of the variable *alive* under the Auction assignment system. Construct a 90% confidence interval for this mean. Give an interpretation of these results in a sentence.

4. Call the difference in *alive* between the Auction and Lottery assignment systems, D. State an estimator $\hat{D}$ for D and use the estimator to compute an estimate of D. Compute a standard error for $\hat{D}$. Derive a 95% confidence interval for D and interpret in a sentence.

5. Is the number of surviving trees (*alive*) under an Auction assignment system statistically different at the 5% level from the number of trees surviving under a lottery assignment? (Recall the 5 step-procedure for hypothesis testing)

6. Consider the hypothesis test from the previous question (#5). Under the null hypothesis, with what probability would you fail to reject the null hypothesis at the 1% significance level but reject the hypothesis at the 5% significance level? Use a picture in the explanation of your answer.

7. In what follows, we attempt to understand the channel by which self-selection operates. We follow the paper by running the following regression, separately for the lottery and auction group.

$$alive = \beta_0 + \beta_1 educ + \beta_2 assets\_sum + \beta_3 trees + \beta_4 trynew + \beta_5 outsiders + \varepsilon$$

   To get at this question:

- Do these variables have the same sign in both regressions?
- Are they more precisely estimated for the auction or the lottery group? Comment on the t-statistics
- Now look at the R^2. Is it larger for the auction group regression or the lottery group regression.

Taken together - what can you conclude about why the estimates are more or less precise in the auction group?

8. Now we want to see how adding covariates on the right hand side of our equation affects the coefficient on the treatment indicator "auction". We run the following two regressions:

$$alive = \beta_0 + \beta_1 auction + \varepsilon \quad (1)$$
$$alive = \beta_0 + \beta_1 auction + \beta_1 income\_ganyu + \varepsilon \quad (2)$$
$$alive = \beta_0 + \beta_1 auction + \beta_1 income\_ganyu + \beta_2 trees + \varepsilon \quad (3)$$
$$alive = \beta_0 + \beta_1 auction + \beta_1 income\_ganyu + \beta_2 trees + \beta_3 hhsize + \beta_4 acres\_total + \beta_5 icraf + \varepsilon \quad (4)$$

Looking both at R^2 and the evolution of $\hat{\beta}_1$ as we add variables, a) comment on which variable matters in explaining the outcome, and which is likely correlated with the variable "auction" (go through equation by equation). b) What does this tell you about the selection process under the auction mechanism?

Exercise 2: Perception on Global Warming

The Public Policy Institute conducts statewide surveys (across California) to collect information about a variety of topics (health, environment, political attitudes, education). In July 2014, PPIC conducted a survey to gather information about the perceived onset of global warming among Californian voters. Question: Do you believe that global warming has already begun

	Number of observations	Yes (percent)
All voters	1340	60
Republicans	391	35
Democrats	582	75
Independents	367	63

Consider first the overall result (all voters). Let p be the fraction of Californian voters that believe that global warming has already begun.

1. Use the survey results to estimate p. Also estimate the standard error of your estimate.
2. Construct a 95% confidence interval for p. Interpret.
3. Construct a 99% confidence interval for p. Is it larger or narrower than the 95% confidence interval? Why?
4. Is there statistical evidence that more than 30% of Republicans believe that global warming has begun (to answer this question, use the 5 steps for hypothesis testing using a 5% significance level)
5. Is there statistical evidence that the Democrats are more likely to believe that global warming has begun than Independents? Explain (to answer this question use the 5 steps for hypothesis testing)