14.771
Problem Set 5 - GMM

This problem set is derived from Alatas et al. (2016), "Self-Targeting: Evidence from a Field Experiment in Indonesia". You will not be able to complete the problem set unless you have read it carefully.

Question 1: General

1. Explain the rational behind using "ordeals", such as a long wait time or far travel distance, as a mechanism for targeting the poor. What are some theoretical reasons for why this might not be an effective means of targeting? (Max 6 lines)

2. From the perspective of the government, through what mechanisms would self-selection potentially affect targeting? Explain how each would impact the profile of beneficiaries relative to automatic screening.

Question 2: Model

For this problem, assume that the all households are unsophisticated (i.e., $\alpha = 0$).

1. Write down and explain in words the household's expected utility function for both possible cases (i.e., applying versus not applying for benefits). What is the implication of the linearity assumption?

2. Assume for now that the application cost function is given by $c(l, y) = \tau l \phi y$. What three assumptions does this capture? (max 2 lines per assumption)

3. Suppose also that there are no household shocks, i.e. $\varepsilon = 0$.

   (a) Write down $h(y, l)$ in this case.

   (b) Define y^* as the value of y such that $h(y, l) = 0$. If $y < y^*$, will households apply? Explain in words the two types of selection operating here.

   (c) How does y^* depend on l?

4. Now suppose $\varepsilon \neq 0$.

   (a) What is the interpretation of ε? What is the point of including it?

(b) Write down $h(y, l)$ in this case.

(c) Derive and interpret equation 12 of the paper.

5. In this model, what are three potential reasons that higher l does not necessarily lead to better targeting? (max 3 lines per reason)

Question 3: Structural Estimation

In this section, you will perform a simplified version of the structural estimation presented in Section 7 of the paper. You are free to use any language you'd like, but the TA will only be able to help you with Matlab or Julia.

We will make the following simplifications in comparison to the paper:

- All households are unsophisticated ($\alpha = 0$).
- Assume linear utility.
- Impose $\lambda(y)$ as the empirical λ, rather than using the equilibrium condition in equation (8).
- Ignore measurement error in incomes.

We have provided you with the following additional data, rather than having you calculate it separately:

- Total application costs. This is the variable "totcost_pc".

You might also want to bear the following in mind:

- If an applicant is approved, they receive the benefit for six years, starting in one year's time. You can assume that they receive the benefit in six equal, annual instalments. The variable "benefit" in the dataset is the monthly benefit from the program (the annual benefit divided by 12).

1. What is the purpose of structurally estimating *this* model?
2. Using the companion data set "data_ps5.csv", generate indicators for each household in terms of (1) quintiles of observed consumption, (2) terciles of PMT scores, (3) terciles of **unobserved** consumption (unobserved consumption = the residual from a regression of log(observed consumption) on the PMT score), and (4) quartiles distance from the registration site.

   Please provide a table with the min, max, and relevant quantiles for each of these variables (by relevant we mean the different quantiles indicated for each variable)
3. For the following groups, calculate the mean show-up moments in the data.

   (a) Each of the five quintiles of observable consumption for the "far" treatment

   (b) Each of the five quintiles of observable consumption for the "close" treatment

   (c) The households who are in both the top tercile observable consumption and bottom tercile unobservable consumption

   (d) The households who are in both the top tercile observable consumption and top tercile unobservable consumption

(e) The households who are in both the bottom tercile observable consumption and bottom tercile unobservable consumption

(f) The households who are in both the bottom tercile observable consumption and top tercile unobservable consumption

(g) Top distance quartile

(h) Bottom distance quartile

4. Given the simplifications above, what is the predicted show-up rate for household i? (Note: This should be a simplified version of equation (22) in the paper).

5. Use GMM (if you are feeling daring you may use a 2-step GMM estimator, but one step is fine as well) to estimate the **two** remaining unknown parameters (given the simplifying assumptions above): the mean and standard deviation of the utility shock ε, ν_ε and σ_ε. (Note: Don't worry about estimating standard errors. If you are feeling double daring though, you may do it by bootstrapping)

14.771: Development Economics

Fall 2021