ECON 272 - 01, 02

Homework 6

Due Date: April 8

Note: Q.1 and Q.2 of this homework uses the dataset *hw7*. This is data on manufacturing firms from 2014. The last page provides the defintions of variables used in the dataset.

Q.1a). Summarize total number of workers (*nototalworker*) and total output (*total_output*) first (i) without weighting by the multiplier (*mult*), and then (ii) after weighting using the multiplier.

Does the sample average change upon including the multiplier weights? Based on the two sample averages, what information can you gain about which type of firms are "over-sampled?"

Hint 1: to obtain the summary statistics using the sample weights, use the following command: sum x [aw = weight].

Hint 2: weighting makes the sample nationally representative. The weights denote the inverse of the sampling probability.

Note: For questions 1b)-1i), always weight your regressions, as well as any summary statistics you compute using *mult*. The regression command is *reg y x [aw = weight].*

Q.1b) We will explore the hypothesis of whether firm profits are affected by the presence of banks. The population regression function is:

$$ShProfit_{id} = \beta_0 + \beta_1 BranchPC_d + \delta \mathbf{X}_{id} + \epsilon_{id}$$

ShProfit is profits earned by firm i, divided by total firm assets. *BranchPC* is bank branches per million population in the district in which firm i is located. $\mathbf{X}$ denotes covariates. Include the following covariates: fixed assets (*avg_nfa*), total workers (*nototalworker*), raw materials (*avg_raw_mat*), a quadratic in age (*age* and *sq_age*) whether the firm is publicly listed (*listed*), and whether the firm is an importer (*importer*).

Report the sign and statistical significance of β_1.

Q.1c) Now re-run the specification in b), but winsorize profits at the **top** and **bottom** 1 percent. How has the β_1 coefficient changed in terms of sign and statistical significance upon winsorizing?

Hint: recall, winsorizing the top 1 percent of a variable means replacing any value in excess of the 99 percentile with the value of the 99 percentile.

Q.1d) Would it be appropriate to use a natural log transformation of the $ShProfit$? How would the regression change if we apply a log transformation? Explain your reasoning.

Q.1e) Set up and test the hypothesis of whether profits are significantly different for firms which are both small (less than 20 workers) and young (less than 10 years of age). These firms are classified by the binary variable *small_young*. Include all covariates from Q.1b) except for *age* and *sq_age*, and also include $BranchPC$ in the estimation. Do small and young firms report significantly different profits than small and old firms?

Q.1f) Now consider the following population regression function:

$$Output_{id} = \beta_0 + \beta_1 BranchPC_d + \delta \mathbf{X}_{id} + \epsilon_{id}$$

Output refers to total manufacturing output (*total_output*). Report the sign of the estimated β_1 coefficient and its statistical significance. Use all the controls specified in Q.1b).

Q.1g) Create a new variable *woutput* which is total output, but winsorized at the top 1%. Re-estimate the population regression function in Q.1f), but using the winsorized outcome variable *woutput*. How does the coefficient on β_1 change in terms of sign and statistical significance, relative to Q.1f)?

Q.1h) Re-estimate the population regression function in Q.1f) but using the natural log of total output as the outcome variable (take natural log of the *non-winsorized* value of output). Based on the estimated β_1, would you say bank branches have a large or small impact on manufacturing output?

Q.1i) Based on the estimates of β_1 in Q.1g) and Q.1h), how different are your results from winsorizing, as opposed to taking the natural log of the outcome variable? When comparing the estimate of β_1 between Q.1f), Q.1g) and Q.1f), what do you think winsorizing does to the regression estimates (consider both the estimated value of β_1 and the standard error of

the β_1 coefficient).

Q.2a) Consider the following population regression function to test whether capital investment responds to bank branches. Remember to weight all your regressions using *mult*.

$$Pr(Capex = 1)_{id} = \beta_0 + \beta_1 BranchPC_d + \delta \mathbf{X}_{id} + \epsilon_{id}$$

$Pr(Capex = 1)$ is a binary equaling 1 if the firm undertook any capital investment (*pcapex*). In $\mathbf{X}$, include the binary variables *listed*, *small*, *young* and *importer*. Does bank branches have a large or small impact on capital investment?

Q.2b) Test whether the impact of bank branches on the likelihood of making any capital investment differs across small and large firms. Estimate the following regression specification:

$$Pr(Capex = 1)_{id} = \beta_0 + \beta_1 BranchPC_d + \beta_2 BranchPC_d \times Small_{id} + \delta \mathbf{X}_{id} + \epsilon_{id}$$

Interpret the β_1 and β_2 coefficients. Can you reject the null hypothesis $H_0 : \beta_1 + \beta_2 = 0$? What information is provided by $\beta_1 + \beta_2$?

Q.3. The government would like to measure how public policies affect firm profits. The specification of interest is: $Profit_{ij} = \beta_0 + \beta_1 Tariff_j + \epsilon_{ij}$. The unit of observation is firm i, operating in industry j. $Tariff_j$ is the average tariff rate imposed on imports on products in industry j. The data on imports is obtained from administrative data, but each firm self-reports profits, which is measured with error.

a) Assuming no other covariates, and assuming $Cov(Tariff_j, \epsilon_{ij}) = 0$, how would measurement error in firm profits affect the estimated β_1 (think of unbiasedness) and the standard error of $\hat{\beta}_1$?

b) Now consider that we add covariates to our regression specification. One of the covariates is $Lobby_j$, which is the total amount spent by lobbyists representing industry j. This variable is measured with error. Would the inclusion of this variable affect the estimated β_1 (again, think of unbiasedness)?

Variable Definitions

Variable names in the Stata dataset are in parentheses

Multiplier (mult): these are the multiplier weights associated with each firm. The weights measure how likely it is to a see this sampled firm in the population distribution. **State** (state): state in which the firm is located.
District (dist01): district in which the firm is located.
Rural (rural): binary variable equaling 1 if the firm is located in a rural area.
Total workers (nototalworker): number of workers hired by the firm.
Total output (total_output): total output produced by the firm in the year (in millions of USD).
Total wages (tot_wage_final): total wages paid by firm to workers (in millions of USD).
Capital (avg_nfa): total capital stock of the firm (in millions of USD).
Assets(avg_assets): total assets of the firm (in millions of USD).
Profits(total_profit): annual profits of the firm (in millions of USD).
Loans(avg_loan): total outstanding loans of the firm (in millions of USD).
Importer (importer): binary variable equaling 1 if the firm imported any input during the year.
Raw Materials (avg_raw_mat): value of raw materials used by the firm during the year (in USD).
Listing status (listed): binary variable equaling 1 if the firm was publicly listed in the stock market.
State-owned firm (psu): binary variable equaling 1 if the firm was owned by the government.
Age (age): age of the firm in years. **Branch per capita**(branch_pc): bank branches in the district where the firm is located, per million persons.
Small Firm(small): binary variable equaling 1 if the firm hires less than 20 workers.
Young Firm(young): binary variable equaling 1 if the firm is aged less than 10 years.
Small and Young Firm(small_young): $Small \times Young$.