Econ 272

Homework 6

Name: Gianluca Crescenzo

Question 1.

(a) Summarize the total number of workers (*nototalworker*) and total ouput (*total_ output*) first without weighting by the multiplier (*mult*), and then 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"?

Solution. Stata gave the following output:

```
. sum nototalworker total_output

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
nototalwor~r |      6,898    290.7061    741.6751          0      21853
total_output |      6,898    121.1889     964.988   .0001354   37129.21

.
. sum nototalworker total_output [aw = mult]

    Variable |     Obs      Weight        Mean   Std. dev.       Min        Max
-------------+-----------------------------------------------------------------
nototalwor~r |   6,898       15698    143.7318   510.0754          0      21853
total_output |   6,898       15698    60.27405   642.4071   .0001354   37129.21
```

We can see that the sample average did change upon including the multiplier weights. Since the average total number of workers *decreased* after including the weights, larger firms were over-sampled.

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

$$\text{ShProfit}_{id} = \beta_0 + \beta_1 \text{BranchPC}_d + \delta \mathbf{X}_{id} + \epsilon_{id}.$$

ShProfit is the profits earned by firm i, divided by the total firm assets. BranchPC is bank branches per million population in the discrict 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.

Solution. Here is my Stata output:

```
------------------------------------------------------------------------------
     sh_proft | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   branch_pc |  -.0001441   .0000911    -1.58   0.114    -.0003227    .0000344
     avg_nfa |  -5.39e-06   .0000264    -0.20   0.838    -.0000571    .0000464
nototalworker |   4.26e-07   9.21e-06     0.05   0.963    -.0000176    .0000185
 avg_raw_mat |   9.18e-06   .0001936     0.05   0.962    -.0003704    .0003888
         age |  -.0017386   .0006052    -2.87   0.004    -.0029249   -.0005523
      sq_age |   .0000185   6.45e-06     2.87   0.004     5.85e-06    .0000311
      listed |  -.0296402   .0154571    -1.92   0.055    -.0599409    .0006604
    importer |   .0336938   .0107916     3.12   0.002      .012539    .0548487
       _cons |    .113839    .013368     8.52   0.000     .0876336    .1400444
------------------------------------------------------------------------------
```

We can see that $\widehat{\beta}_0 = -0.0001441$. With a p-value of 0.114, our value is not statistically significant at the 5% level.

(c) Now re-run the specification in (b), but winsorize profits at the top and bottom 1%. How has the β_1 coefficient changes in terms of sign and statistical significant upon winsorizing?

Solution. Observe that:

```
. winsor sh_profit, p(0.01) generate(shprofit_w)

.
. reg shprofit_w branch_pc avg_nfa nototalworker avg_raw_mat age sq_age listed imp
> orter [aw = mult]
(sum of wgt is 15,665)

.
```

```
------------------------------------------------------------------------------
  shprofit_w | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   branch_pc |  -.0000955   .0000505    -1.89   0.059    -.0001944    3.44e-06
     avg_nfa |  -3.16e-06   .0000146    -0.22   0.829    -.0000318    .0000255
nototalworker |   2.45e-06   5.10e-06     0.48   0.631    -7.56e-06    .0000125
 avg_raw_mat |  -7.05e-06   .0001073    -0.07   0.948    -.0002173    .0002032
         age |  -.0010709   .0003353    -3.19   0.001    -.0017282   -.0004137
      sq_age |   .0000104   3.57e-06     2.91   0.004     3.39e-06    .0000174
      listed |  -.0276464   .0085633    -3.23   0.001     -.044433   -.0108597
    importer |   .0332783   .0059786     5.57   0.000     .0215585    .0449982
       _cons |   .0897713   .0074059    12.12   0.000     .0752534    .1042892
------------------------------------------------------------------------------
```

The sign has not changed. Although the coefficient is still not statistically significant, it is much closer than before.

(d) Would it be appropriate to use a natural log transformation of the ShProfit? How would the regression change if we apply a log transformation.

Solution. If we were to consider doing a log-transform, first note that the β_1 coefficient represents a percentage effect rather than an absolute-change effect. Since ShProfit is a ratio, it might be beneficial to consider it's log-transform, but only if all the values are strictly positive.

(e) 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 (b) 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?

Solution. From the table:

sh_profit	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
branch_pc	-.0001615	.0000909	-1.78	0.076	-.0003398	.0000167
avg_nfa	-4.30e-06	.0000264	-0.16	0.871	-.0000561	.0000475
nototalworker	6.76e-07	9.17e-06	0.07	0.941	-.0000173	.0000187
avg_raw_mat	2.96e-06	.0001937	0.02	0.988	-.0003767	.0003826
listed	-.0321119	.0154248	-2.08	0.037	-.0623494	-.0018745
importer	.0333153	.0107932	3.09	0.002	.0121572	.0544733
small_young	-.0341555	.0193569	-1.76	0.078	-.0721009	.00379
_cons	.0894562	.0092007	9.72	0.000	.07142	.1074924

We can see there is a -0.034 difference in ShProfit. This is not statistically significant at the 5% level.

(f) Now consider the following population regression function:

$$\text{Output}_{id} = \beta_0 + \beta_1 \text{BranchPC}_d + \delta \mathbf{X}_{id} + \epsilon_{id}.$$

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

Solution. Running the regression gave the following table:

total_output	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
branch_pc	.0276864	.1098674	0.25	0.801	-.1876877	.2430606
avg_nfa	.0136459	.0318454	0.43	0.668	-.0487809	.0760727
nototalworker	-.00101	.0111142	-0.09	0.928	-.0227972	.0207773
avg_raw_mat	16.09031	.2335668	68.89	0.000	15.63245	16.54818
age	.6200993	.7299844	0.85	0.396	-.8108959	2.051095
sq_age	-.0044073	.0077835	-0.57	0.571	-.0196654	.0108508
listed	19.3078	18.64495	1.04	0.300	-17.24207	55.85768
importer	-19.76467	13.01727	-1.52	0.129	-45.28254	5.7532
_cons	-18.49531	16.12504	-1.15	0.251	-50.10538	13.11475

The sign of β_1 is positive, but it is not statistically significant at the 5% or 10% levels.

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

Solution. Stata outputted:

woutput	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
branch_pc	-.1082417	.0280508	-3.86	0.000	-.16323	-.0532535
avg_nfa	-.101354	.0081306	-12.47	0.000	-.1172925	-.0854155
nototalworker	.0891258	.0028376	31.41	0.000	.0835632	.0946884
avg_raw_mat	1.785863	.0596331	29.95	0.000	1.668963	1.902762
age	.419517	.1863759	2.25	0.024	.0541625	.7848715
sq_age	-.0049028	.0019873	-2.47	0.014	-.0087984	-.0010072
listed	81.48681	4.760335	17.12	0.000	72.15508	90.81854
importer	51.66862	3.323503	15.55	0.000	45.15353	58.18372
_cons	12.30627	4.116964	2.99	0.003	4.235749	20.37679

After winsorizing, $\widehat{\beta_1}$ is now negative and statistically significant.

(h) Re-estimate the population regression function in (f) but using the natural log of total output as the outcome variable (take the 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?

Solution. From the table:

ln_total_ou~t	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
branch_pc	-.0024923	.0004974	-5.01	0.000	-.0034673	-.0015172
avg_nfa	-.0016166	.0001442	-11.21	0.000	-.0018992	-.001334
nototalworker	.0011478	.0000503	22.81	0.000	.0010492	.0012464
avg_raw_mat	.0122164	.0010574	11.55	0.000	.0101436	.0142892
age	-.0151372	.0033047	-4.58	0.000	-.0216155	-.0086589
sq_age	.0001422	.0000352	4.04	0.000	.0000732	.0002113
listed	1.541427	.0844081	18.26	0.000	1.375961	1.706893
importer	1.702065	.0589308	28.88	0.000	1.586542	1.817588
_cons	1.635389	.0730001	22.40	0.000	1.492286	1.778492

we can see that bank branches have a small and negative impact on manufacturing output based on the estimated coefficient. A 1-unit increase in BranchPC; i.e., one additional bank branch per million population, results in a change of -0.24923% to total output.

(i) Based on the estimates of β_1 in (g) and (h), 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 (f), (g), and (h), 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.)

Solution. Both (g) and (h) resulted in β_1 coefficients which were statistically significant, whereas the coefficient in (f) was not statistically significant. Comparing winsorizing and taking the natural log of the outcome variable, there was a large change in the standard error.

Question 2.

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

$$P(\text{Capex} = 1)_i d = \beta_0 + \beta_1 \text{BranchPC}_d + \delta \mathbf{X}_d + \epsilon_{id}.$$

$P(\text{Capex} = 1)$ is a binary equaling to 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?

Solution. From the table:

pcapex	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
branch_pc	.0001537	.0001338	1.15	0.251	-.0001086	.000416
listed	-.0586969	.0226346	-2.59	0.010	-.1030678	-.0143261
small	-.1291336	.0120588	-10.71	0.000	-.1527726	-.1054945
young	-.0320557	.0205018	-1.56	0.118	-.0722455	.0081341
importer	.0769497	.0160723	4.79	0.000	.0454431	.1084563
_cons	.3688157	.0148515	24.83	0.000	.3397022	.3979293

we can see that bank branches has a small impact on capital investment.

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

$$P(\text{Capex} = 1)_i d = \beta_0 + \beta_1 \text{BranchPC}_d + \beta_2 (\text{BranchPC}_d \cdot \text{Small}_{id}) + \delta \mathbf{X}_d + \epsilon_{id}.$$

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 $\beta_1 + \beta_2$.

Solution. Stata gave the following table:

pcapex	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
branch_pc	.0003948	.0001919	2.06	0.040	.0000187	.0007709
branch_small	-.0004689	.0002674	-1.75	0.080	-.0009931	.0000553
listed	-.0596855	.0226382	-2.64	0.008	-.1040634	-.0153076
small	-.0875657	.0265934	-3.29	0.001	-.1396971	-.0354344
young	-.0323447	.0204993	-1.58	0.115	-.0725298	.0078403
importer	.0771489	.0160702	4.80	0.000	.0456462	.1086515
_cons	.3479739	.0190196	18.30	0.000	.3106897	.3852581

β_1 is the effect of BranchPC on the probability of capital investment for large firms. β_2 tells us how the effect of BranchPC differs for small firms, relative to large firms. Running the test command in Stata outputted:

```
 ( 1)  branch_pc + branch_small = 0

       F(  1,  6874) =    0.16
            Prob > F =    0.6912
```

So we cannot reject the null-hypothesis. $\beta_1 + \beta_2$ is the total effect for small firms.

Question 3.

(a)

(b)