Applied Microeconometrics I
Discussion of Problem Set 2

Atte Pudas

Aalto University

27/09/2023

Feedback

- Average grade 74/100

- Respect the word limits: clear violations result in loss of points
- Late submissions are not accepted. No exceptions.
- Provide the log file. You need to show relevant code and output to get full points from Stata questions.

"Tips":

- Answer to what is asked and try to be specific.
- When you write something to your answers, clearly tell what do you mean, and to which part of the question you are answering. For example, if you are asked "What is the main contribution?", it's recommended that you answer something like "The main contribution is ..."
- Think what is important, do not focus on details.
- Concise answers are expected, go straight to the point.

Exercise 1

General equilibrium effects in randomized experiments: The case of displacement effects of labor market policies, Cheung et al. (2023)

Exercise 1.1

Briefly summarize the paper, explaining what its contribution to the literature is. Why is focusing on general equilibrium effects important?

- The paper uses two-level randomization and estimates a direct positive effect of job search assistance (JSA) on the exit rate from unemployment for the treated, and a negative displacement effect (GE effect) for the untreated.
- Main contribution is to show why the JSA program helps the treated job seekers (benefited from better information on relevant open jobs) and the types of displacement effects at work (more competition for jobs, not reallocation of resources).
- A standard experiment is not enough to deal with GE effects. Ignoring GE effects prevents from observing the net effect of the experiment. Need for two-level randomization.

Exercise 1.2

How does the identification strategy adopted by the researchers help to alleviate the general equilibrium effect problem in this example?

- GE/spillover effects should happen only in labor markets where a large share of individuals are treated.
- The authors apply a two-level randomization strategy over both offices and job seekers (within active offices). Randomization over offices identifies the displacement effects by comparing non-treated job seekers at active (treated) and non-active (control) offices.

  (Many local labor markets in Sweden have one employment office, so randomization over offices in many cases implies randomization over local labor markets.)

Exercise 1.3

What are the "active" and "non-active offices" in the paper? How is individual-level treatment status assigned?

- Treated and control offices. Randomization over job seekers occurs at the active offices.
- Treatment status is randomly assigned to individuals via date of birth.

Exercise 1.4

How is the direct effect of the program identified and estimated? Comment on the statistical significance and economic magnitude of the direct effect estimates in Table 3

- Randomization over job seekers within active (treated) offices identifies the direct effect of the JSA program by comparing treated and non-treated job seekers in the same labor market.
- Estimated with model (1), Coef. of interest is β_0 (also with model (2), β_1):

$$Y_i = \alpha_0 + \beta_0 1(\text{Assigned to program}_i) + \rho X_i + \varepsilon_i,$$

- The coefficients in row 1 in Table 3 imply a statistically significant positive direct effect: the JSA program (treatment) increases the probability of leaving unemployment during the first quarter of unemployment by 3.5 percentage points, or about 10%.

Exercise 1.4

Table 3: Direct and displacement effects (ITT) of the JSA program on the exit from unemployment during the first quarter

	Experiment period in 2015			2012–2015	
	(1)	(2)	(3)	(4)	(5)
Assigned to program	0.036***	0.035***	0.035***	0.035***	0.034***
	(0.006)	(0.006)	(0.005)	(0.006)	(0.006)
In a program area			−0.015	−0.016**	−0.015**
			(0.010)	(0.007)	(0.006)
Net effect treated			0.020**	0.019***	0.018***
			(0.009)	(0.006)	(0.005)
Control mean	0.354	0.354	0.368	0.390	0.390
Year dummies	No	No	No	Yes	Yes
Month dummies	No	No	No	Yes	Yes
PES office dummies	No	No	No	Yes	Yes
Controls	No	Yes	Yes	No	Yes
Clusters	No	No	72	72	72
Observations	26,538	26,538	57,778	552,816	552,816

Exercise 1.5

How are the general equilibrium effects of the program identified and estimated? Comment on the statistical significance and economic magnitude of the general equilibrium estimates in Table 3

- The displacement effect is identified through the comparison of non-treated job seekers at active and non-active offices.
- Estimated with model (2), Coef. of interest is β_2:

$$Y_{ij} = \alpha_1 + \beta_1 1(\text{Assigned to program}_{ij}) + \beta_2 1(\text{In a program area}_j) + \rho X_i + \varepsilon_{ij},$$

- The coefficients in row 2 in Table 3 indicate that the JSA program reduces the exit rate from unemployment for the non-treated job seekers at the active offices by 1.5 percentage points (3.8%) during the first quarter, suggesting substantial displacement effects. The precision of the estimates is low without including exogenous covariates.

Exercise 1.5

Table 3: Direct and displacement effects (ITT) of the JSA program on the exit from unemployment during the first quarter

	Experiment period in 2015			2012–2015	
	(1)	(2)	(3)	(4)	(5)
Assigned to program	0.036***	0.035***	0.035***	0.035***	0.034***
	(0.006)	(0.006)	(0.005)	(0.006)	(0.006)
In a program area			−0.015	−0.016**	−0.015**
			(0.010)	(0.007)	(0.006)
Net effect treated			0.020**	0.019***	0.018***
			(0.009)	(0.006)	(0.005)
Control mean	0.354	0.354	0.368	0.390	0.390
Year dummies	No	No	No	Yes	Yes
Month dummies	No	No	No	Yes	Yes
PES office dummies	No	No	No	Yes	Yes
Controls	No	Yes	Yes	No	Yes
Clusters	No	No	72	72	72
Observations	26,538	26,538	57,778	552,816	552,816

Exercise 2

DiPasquale and Glaeser (1999) find, using survey data from the US, that homeowners (right-hand side variable) make better citizens, as measured by self-reported measures such as the number of non-professional organizations an individual belongs to, voting in local elections, church attendance and gardening (left-hand side variables).

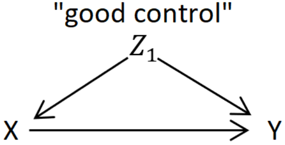
Exercise 2.1

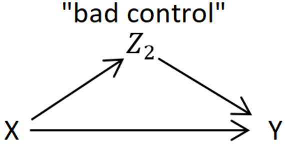
Explain verbally under which assumptions we can interpret this correlation as the causal impact of homeownership on citizenship.

- Conditional on controls that the authors include in the regressions, homeownership is as good as randomly assigned (conditional independence assumption = CIA)

Good and bad controls

- Examples of a good and a bad control Z_1 and Z_2:





- "Don't run a regression like
  $wage = a + b\ education + c\ industry + error$.
  Of course, adding industry helps raise the R^2, and industry is an important other determinant of wage. But the whole point of getting an education is to help people move to better industries, not to move from assistant burger-flipper to chief burger-flipper."[1]
- You can check "A Crash Course in Good and Bad Controls" by Cinelli et al. (2022) for an interesting discussion about good and bad controls.

[1] Cochrane, J. H. (2005). Writing tips for Ph. D. students (pp. 1-12). Chicago, IL: University of Chicago.

Exercise 2.2

The authors show that this correlation is robust to the inclusion as controls of family status. Discuss whether this is a "good" control.

- Potentially a "bad" control if homeownership affects the outcomes of interest through family status (e.g. marriage).
- However, family status may affect both homeownership and the outcomes of interest.

Exercise 2.3

The authors also control for individuals' education level. Discuss whether this is a "good" control.

- Potentially a "bad" control if homeownership affects outcomes of interest through educational outcomes.
- However, education may affect both homeownership and the outcomes of interest.

Exercise 2.4

The left-hand side variables are self-reported. Discuss whether this might cause measurement error and explain how this affects the interpretation of coefficients.

- Self-reported values may suffer from measurement error: for example, individuals may not remember things correctly or may not answer honestly.
- Random error in LHS variable will lead to imprecision (higher standard errors), but our estimates will still be consistent.
- Systematic error in LHS variable will lead to biased estimates.

Exercise 2.5

Control variables are derived from survey data. Discuss how this might cause measurement error in control variables and how this affects the interpretation of results.

- Random error on RHS variables will cause attenuation bias i.e. coefficients are biased towards zero.
- Thus, we will underestimate the sensitivity of the coefficient of interest to controls.
- This makes it more difficult to use controls to infer the extent of potential omitted variable bias.

Exercise 3

Do Employers Use Unemployment as a Sorting Criterion When Hiring? Evidence from a Field Experiment", Eriksson and Rooth (2014)

Exercise 3.1

Why is it difficult to study the role of past unemployment in employers' hiring decisions?

- Unobservables: "While it is clear that the exit rate from unemployment to work declines with the length of the spell in most countries ... **it is difficult to separate the effects of unemployment from the effects of other important worker characteristics which are observed by the recruiting firms but not by the researcher**."

Exercise 3.2

How do the authors estimate the causal impact of unemployment and employment history? Explain briefly their empirical strategy.

- The authors randomly sent fictitious job applications to employers.
- Information about the workers' employment history were **randomly** assigned to the applications.
- Since employers invite job seekers to interview based only on their application, the authors can isolate the causal effect of the randomly assigned employment history: there are no inter-dependencies among the regressors, and there is no scope for unobserved heterogeneity with respect to worker characteristics.

Exercise 3.3

According to you, what are the limitations of this study? Mention at least two of them.

For example:

- It is not possible to distinguish the effect of past labor market history from that of age.
- The authors study only the early stages of the hiring process. They do not know whom the employers eventually decided to hire.

Exercise 3.4

Estimate the effect of years of work experience on the probability of being called back for all jobs and by type of job (see Table 9, panel A). Report the estimated coefficients and standard errors. Discuss what these estimates mean and explain the magnitude of the effect in economic and statistical terms. Based on these estimates, does the impact of the treatment differ between types of jobs in a meaningful way?

Exercise 3.4

- Run the following commands:

```
* Table 9, Panel A

* All jobs
dprobit callback erfar arblosnu_3 arblosnu_6 arblosnu_9 scarce parbetslos antlarbetsg min kvinna ///
doverutb sommarjobb utlgymn komp varme cult fotboll basket lopning simning golf tennis dgbg dlandet ddata dmano dsvso dgym dred dsjuk ///
dbutik dftg drest dlokal dmaskin dbygg dfordon ///
, cluster (id)

*Medium/low skill jobs
dprobit callback erfar arblosnu_3 arblosnu_6 arblosnu_9 scarce parbetslos antlarbetsg min kvinna ///
doverutb sommarjobb utlgymn komp varme cult fotboll basket lopning simning golf tennis dgbg dlandet ddata dmano dsvso dgym dred dsjuk ///
dbutik dftg drest dlokal dmaskin dbygg dfordon ///
if dhighskill==0, cluster (id)

*High skill jobs
* this is compared with results at question 3.5
dprobit callback erfar arblosnu_3 arblosnu_6 arblosnu_9 scarce parbetslos antlarbetsg min kvinna ///
doverutb sommarjobb utlgymn komp varme cult fotboll basket lopning simning golf tennis dgbg dlandet ddata dmano dsvso dgym dred dsjuk ///
dbutik dftg drest dlokal dmaskin dbygg dfordon ///
if dhighskill==1, cluster (id)
```

- The three regressions give you the coefficients (marginal effects) for variable "erfar" (work experience in years):

Table 9—The Effect of the Workers' Work Experience on the Callback Rate (*Marginal Effects*)

	All jobs	Medium/ low skill jobs	High skill jobs
Panel A			
Work experience (in years)	0.012***	0.009*	0.017**
	(0.004)	(0.005)	(0.008)

Exercise 3.4

- Statistically significant positive effects: the effect of one additional year of experience on callback rate was 1.2, 0.9, and 1.7 percentage point for all, medium/low, and high skill jobs, respectively.
- The effects are economically meaningful, they translate to about 5-6% increases compared to the average callback rates. The differences in the estimates suggest that work experience matters differently for medium/low and high skill jobs, even though the 95% confidence intervals are overlapping.

Exercise 3.5

Suppose now that you are interested in how work experience affects the callback rate for high-skilled workers, and that you estimate a model where you regress the outcome on experience in year (without including all other regressors). How does the estimate and its significance compare to that at the previous point? Interpret the result. Is your answer different when you specify experience with the "2 year" and "3-5 years" dummies?

Exercise 3.5

- For linear experience, run command

```
* high-skilled, no controls, linear experience
dprobit callback erfar if dhighskill==1, cluster (id)
```

You'll get

callback	dF/dx	Robust Std. err.	z	P>|z|	x-bar	[95%	C.I.]
erfar	.01036	.0065075	1.59	0.112	3.0285	-.002394	.023114

- The coefficient is different from the 0.017 above and not statistically significant.

Exercise 3.5

- For categorized experience, run command

```
* high-skilled, no controls, dummies for experience
dprobit callback erfar2 erfar3_5 if dhighskill==1, cluster (id)
```

You'll get

callback	dF/dx	Robust Std. err.	z	P>|z|	x-bar	[95% C.I.	]
erfar2*	.0540091	.0296066	1.86	0.063	.195377	-.004019	.112037
erfar3_5*	.057994	.0234737	2.43	0.015	.661495	.011986	.104002

- Note that we must leave category "experience of 1 year" as the base group to which marginal effects of other categories are compared to
- The coefficients are different from the corresponding coefficients in Table 3 (0.079 and 0.088 for the categories of 2 and 3-5 years of experience, respectively) and also statistically less significant.

Exercise 3.5

- The estimates are different because we omit labor market regressors that are correlated with work experience by design:

  If application was given one employer, years of work experience was randomly given a value between 1 and 5. If given three employers, values between 3 and 5 were given. This means that number of employers and work experience are, by construction, correlated. However, conditional on number of employers (and unemployement between jobs which is also correlated with number of employers by design), years of work experience is independent.
- Exclusion of other variables (that should be independent from work experience) reduces precision but (in large samples) should not bias the estimate.

Exercise 3.6

See Section Robustness at the end of the paper. Estimate a model to verify whether "unemployment after graduation" is correlated with the worker characteristics that are not employment history variables. Test for the joint significance of the model coefficients and interpret the result. Is your answer different when on the right-hand side we also include all the employment variables that are not "unemployment after graduation"? Interpret the result.

Exercise 3.6

- Run command:

```
dprobit scarce min kvinna doverutb sommarjobb utlgymn komp varme cult ///
   fotboll basket lopning simning golf tennis dgbg dlandet ddata dmano dsvso dgym dred dsjuk ///
   dbutik dftg drest dlokal dmaskin dbygg dfordon, cluster (id)
```

- You'll find the test for joint significance on top of the regression output:

```
Probit regression, reporting marginal effects           Number of obs =    8466
                                                        Wald chi2(28) =   16.97
                                                        Prob > chi2   =  0.9491
Log pseudolikelihood = -4192.8497                       Pseudo R2     =  0.0020

                                 (Std. err. adjusted for 3,821 clusters in id)
```

scarce	dF/dx	Robust Std. err.	z	P>|z|	x-bar	[95% C.I.	]
min*	.0003992	.0107451	0.04	0.970	.331207	-.020661	.021459
kvinna*	.0196508	.0105669	1.88	0.060	.334278	-.00106	.040362
doverutb*	-.0083985	.010042	-0.83	0.406	.242381	-.028081	.011283
sommar~b*	.0094937	.0088643	1.07	0.283	.398535	-.00788	.026868
utlgymn*	-.0046415	.0172491	-0.27	0.789	.071935	-.038449	.029166

- We do not reject that all coefficients equal zero. This is expected since unemployment after graduation should be independent from these covariates (all worker characteristics should be uncorrelated with the employment history variables)

Exercise 3.6

- Add the employment variables and run command:

```
dprobit scarce arblosnu_3 arblosnu_6 arblosnu_9  parbetslos erfar antlarbetsg ///
min kvinna doverutb sommarjobb utlgymn komp varme cult ///
fotboll basket lopning simning golf tennis dgbg dlandet ddata dmano dsvso dgym dred dsjuk ///
dbutik dftg drest dlokal dmaskin dbygg dfordon, cluster (id)
```

- Now the joint test is almost significant with 5% level:

```
Probit regression, reporting marginal effects           Number of obs =    8466
                                                        Wald chi2(34) =   48.54
                                                        Prob > chi2   =  0.0506
Log pseudolikelihood = -4177.2811                       Pseudo R2     =  0.0057
```

- This is because the added variables are in general correlated with the "unemployment after graduation" by construction.