University of Zurich **Spring term 2019**

Department of Economics

Econometrics for Research Students II
Problem Set #5

(due on Friday, April 19, 2019)

Rules:

- Solution must be submitted on GitHub on the following repository econometrics-for-research-students-ii/Problem-set-5 by **12 am** (midnight). Late assignments will not be graded. If you don't have an GitHub account yet, please sign up at this link.
- Please append your STATA (or equivalent program) "do" files to your solutions.
- Everybody has to submit an own solution.
- Please indicate the members of your 'discussion group' on the assignment, if the solution was discussed prior to submission.

The Roy Model

5.1 Generalized Roy Model: The union/non-union wage differential

This exercise is based on the seminal article

Lung-Fei Lee (1978): "Unionism and wage rates: A simultaneous equations model with qualitative and limited dependent variables", *International Economic Review*, **19**: 415-433.

Based on the idea that "[e]conomic considerations suggest that the propensity to join a union depends on the net wage gains that might result from trade union membership", Lee (1978) proposed the estimation of what is now known as a (parametric) switching regression model. Here, we consider a simplified version of his model. Assume that every worker has two different potential wages, depending on his union membership status

$$\begin{aligned} \ln w_i^U &= x_i'\beta^U + u_i^U, \quad u_i^U \sim Normal(0, \sigma_U^2) \\ \ln w_i^N &= x_i'\beta^N + u_i^N, \quad u_i^N \sim Normal(0, \sigma_N^2) \end{aligned}$$

Where U stands for unionized, and N for non-unionized. A worker decides to join the union if

$$U_i^* = \delta_0 + \delta_1(\ln w_i^U - \ln w_i^N) + x_i'\delta_2 + z_i'\delta_3 - v_i > 0$$

and v_i is assumed to be distributed as $Normal(0, \sigma_v^2)$. For simplicity, assume that all the error terms are uncorrelated with each other, $Cov(\sigma_U^2, \sigma_N^2) = Cov(\sigma_U^2, \sigma_v^2) = Cov(\sigma_N^2, \sigma_v^2) = 0$. For every worker, only $(\ln w_i, x_i, z_i, d_i)$ is observed, where d_i is a union membership indicator equal to 1 if i is a member and 0 else; and $\ln w_i = d_i \ln w_i^U + (1 - d_i) \ln w_i^N$

a. What is the expected union/non-union wage differential for a randomly chosen individual with characteristics x_i in this model? What is the expected wage differential for a union worker with traits x_i?

   Hint: Normalize the selection equation to be a standard probit.

b. Is it possible to obtain an unbiased estimate of β^U and β^N by running two linear regressions using the sample of union members and non-members, respectively?

c. Sketch how both equations can be estimated jointly by maximum likelihood.

   Hint: Use the same specification for the reduced form error as Lee.

d. Describe a two-step method which estimates β^U and β^N consistently by including estimated sample-selection correction variables in the structural wage equations.

e. How can the structural parameters of the union status equation, i.e. δ, be recovered after this estimation procedure?

Now consider Tables 1, 2, 6, and 7 from Lee (1978), which contain the estimation results.

f. Is the **relative** rate of return to education (ED ranges from 1 to 5, where 5 is the highest education level) higher in the unionized or the non-unionized sector?

   What about the effects of market experience (ME, where ME_2 is the square of labor market experience), female ($sex = 0$), blacks ($race = 0$), and health impediments (HLT) on wages?

g. Interpret the sign of the selectivity variables in Tables 1 and 2.

   Hint: Check how Lee (1978) defines the inverse Mills ratio.

h. Judging from the presented evidence, how important would you say is the wage differential in explaining the probability of union membership?

i. Compare Table 6 and Table 7. How can the reduced form estimates be interpreted?

TABLE 1[4]

THE UNION WAGE EQUATION ESTIMATES (SELECTIVITY BIAS ADJUSTED)

Exogenous Variable	Coefficients	T-Values
Constant	4.431	27.129 **
N. E.	−0.083	−3.369 **
N. C.	−0.007	−0.240
S	−0.172	−5.422 **
UR$_1$	0.067	3.279 **
UR$_2$	−0.092	−3.667 **
ED$_1$	−0.108	−2.666 **
ED$_2$	−0.033	−1.330 Δ
ED$_3$	0.052	2.60 **
ED$_4$	0.111	5.168 **
ED$_5$	0.139	4.112 **
ME	0.016	7.526 **
ME$_2$	−0.0002	−5.418 **
RACE	0.095	6.367 **
SEX	0.317	14.915 **
IND$_1$	0.223	4.034 **
IND$_2$	0.169	3.722 **
IND$_3$	0.034	1.477 Δ
IND$_4$	0.018	0.722
U	0.662	6.168 **
HLT	−0.055	−2.105 **
Selectivity Variable	−0.168	−1.914 *

TABLE 2

THE NONUNION WAGE EQUATION ESTIMATES (SELECTIVITY BIAS ADJUSTED)

Exogenous Variable	Coefficients	T-Values
Constant	4.754	71.220 **
N. E.	−0.091	−2.975 **
N. C.	−0.074	−2.238 **
S	−0.139	−4.427 **
UR$_1$	0.039	1.610 Δ
UR$_2$	−0.067	−2.933 **
ED$_1$	−0.049	−1.187
ED$_2$	−0.016	−0.526
ED$_3$	0.087	3.380 **
ED$_4$	0.157	5.979 **
ED$_5$	0.282	6.747 **
ME	0.012	5.468 **
ME$_2$	−0.0002	−4.788 **
RACE	0.186	10.205 **
SEX	0.267	13.501 **
IND$_1$	0.120	1.915 *
IND$_2$	0.130	2.433 **
IND$_3$	0.053	1.474 Δ
IND$_4$	0.058	1.594 Δ
HLT	−0.088	−2.961 **
Selectivity Variable	0.136	3.152 **

TABLE 6

THE UNION STATUS EQUATION ESTIMATES (THE STRUCTURAL FORM ESTIMATES)

	Max. Likelihood Est. Coefficient	Standard Error	
Constant	−0.654	0.145	**
N. E.	0.227	0.076	**
N. C.	0.197	0.077	**
S	−0.296	0.077	**
UR_1	0.129	0.063	**
UR_2	−0.174	0.067	**
ED_1	−0.269	0.116	**
ED_2	−0.098	0.084	
ED_3	0.079	0.072	
ED_4	0.119	0.074	Δ
ED_5	0.258	0.119	**
ME	0.0020	0.0022	
RACE	0.166	0.054	**
SEX	0.093	0.055	*
WK_1	−0.372	0.115	**
WK_2	−0.017	0.073	
CCR	0.365	0.132	**
HLT	−0.185	0.087	**
$\ln \hat{W}_u - \ln \hat{W}_n$	2.455	0.205	**

TABLE 7

THE REDUCED FORM ESTIMATES OF THE UNION STATUS EQUATION

Exogeneous Var.	Coefficients	Standard Error	
Constant	−1.633	0.202	**
N. E.	0.242	0.076	**
N. C.	0.364	0.077	**
S	−0.398	0.077	**
UR_1	0.204	0.063	**
UR_2	−0.240	0.067	**
ED_1	−0.396	0.117	**
ED_2	−0.147	0.084	*
ED_3	−0.007	0.072	
ED_4	0.0024	0.074	
ED_5	−0.095	0.115	
ME	0.013	0.0065	**
ME^2	−0.00004	0.00013	
RACE	−0.061	0.052	
SEX	0.251	0.054	**
WK_1	−0.378	0.115	**
WK_2	−0.029	0.073	
IND_1	0.516	0.184	**
IND_2	−0.115	0.151	
IND_3	−0.098	0.088	
IND_4	0.030	0.095	
U	1.793	0.145	**
CCR	0.433	0.142	**
HLT	−0.101	0.088	

5.2 Application: Female labor supply using Altonji-Elder-Taber Bounds

Let's revisit the Panel Study of Income Dynamics from the 1987 study by Thomas Mroz:

> T.A. Mroz (1987): "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumption", *Econometrica*, **55**: 765-799.

The data set, `mroz.dta`, is available on OLAT. In the following, we want to assess the selection into the labor force `inlf` and motherhood `kidslt6` (generate a dummy variable capturing if a women has more than one child below the age of 6).

a. Run two separate probit regression of `inlf` (and `kidslt6`) on `nwifeinc educ exper expersq age`.

b. Use the `biprobit` command to estimate the probit regressions jointly. What is the advantage of estimating the models jointly?

c. Now suppose that selection into the labor force depends on the selection into motherhood. Where should you include `kidslt6` in your bivariate probit model? What assumptions are you making? What is the estimated correlation ρ between the error terms? Interpret.

In the paper:

> Joseph G. Altonji, Todd E. Elder, and Christopher R. Taber (2008): "Using Selection on Observed Variables to Assess Bias from Unobservables When Evaluating Swan-Ganz Catheterization", *American Economic Review P&P*, **98**(2): 345-350.

the authors develop a new approach to bound a treatment effect accounting for selection on unobservables.

d. Based on their reasoning, explain briefly how to assess the selection on unobservables.

e. Find the value of ρ that eliminates the effect of more than one child on labor force participation. *Hint: Use the* `constraint` *command if you use Stata.*

f. In light of this article, is this value plausible? To answer this question, relate selection on unobservables to selection on observables.

g. Using your estimate from f., compute the lower bound effect of having more than one child on labor force participation.