

Problem Set 6 (SOLUTIONS)

The purpose of the first part of this problem set is to estimate, interpret the results, and compare the results across different binary dependent variable models. In the second part you will estimate and compare different specifications of an endogenous selection model.

First part will be discussed in week 9 and the second part in week 10 of this term.

The data file for this exercise is on Moodle: `mus16data.dta`. It is a subset of the data used by P. Deb, M. Munkin and P.K. Trivedi (2006): “Bayesian Analysis of Two-Part Model with Endogeneity”, *Journal of Applied Econometrics*, 21, 1081-1100. Data is for 2001 and comes from the Medical Expenditure Survey. Sample has 3,328 observations.

The main outcome variable of interest is ambulatory expenditure (`ambexp`) and the regressors are given below.

Since the expenditure data is skewed, we will be using the logged expenditure variable as our dependent variable. You should read Cameron A.C. and Trivedi, P.K. *Micro-econometrics using Stata* to see the pros and cons regarding whether to log the dependent variable or not.

Note, there is one individual who has an `expenditure=1` and this will get coded as 0 when variable is logged. Since it is only one individual, we will ignore the problem by not doing anything. If there are many individuals like this, you will need to see whether you can say why this might be the case.

Dependent variable

- `ambexp`: Ambulatory medical expenditures (excluding dental and outpatient mental). There are 526 individuals with zero expenditure. There is one individual who has expenditure=\$1. I am going to assume that this individual did not spend any money.
- `lambexp`: $\ln(\text{ambexp})$ given `ambexp` > 0 ; missing otherwise
- `dambexp`: 1 if `ambexp` > 0 and 0 otherwise (binary indicator)

Regressors

- `ins`: health insurance measures, either PPO or HMO type insurance
- `totchr`: health status measures: number of chronic diseases
- `age`: age in years/10
- `female`: 1 for females, zero otherwise
- `educ`: years of schooling of decision maker
- `blhisp`: either black or Hispanic
- `income`: income in USD/1000

Preamble

<IPython.core.display.HTML object>

Create a do-file for this problem set and include a preamble that sets the directory and opens the data. For example,

```
clear
//or, to remove all stored values (including macros, matrices, scalars, etc.)
*clear all

* Replace $rootdir with the relevant path to on your local haddrive.
cd "$rootdir/problem-sets/ps-6"

cap log close
log using problem-set-6-log.txt, replace

use mus16data.dta, clear
```

```
C:\Users\neil_\OneDrive - University of Warwick\Documents\EC910\website\warwick
> -ec910\problem-sets\ps-6
```

```
-----
      name:  <unnamed>
      log:   C:\Users\neil_\OneDrive - University of Warwick\Documents\EC910\we
> bsite\warwick-ec910\problem-sets\ps-6\problem-set-6-log.txt
      log type:  smcl
      opened on:  19 Nov 2024, 17:41:06
```

Questions

Part 1

1.1. Obtain and comment on the descriptive statistics for ambexp, lambexp, age, female, educ, blhisp, totchr, ins, income.

```
su dambexp ambexp lambexp age female educ blhisp totchr ins income
```

Variable	Obs	Mean	Std. dev.	Min	Max
-----+-----					

dambexp		3,328	.8419471	.3648454	0	1
ambexp		3,328	1386.519	2530.406	0	49960
lambexp		2,802	6.555066	1.41073	0	10.81898
age		3,328	4.056881	1.121212	2.1	6.4
female		3,328	.5084135	.5000043	0	1
-----+-----						
educ		3,328	13.40565	2.574199	0	17
blhisp		3,328	.3085938	.4619824	0	1
totchr		3,328	.4831731	.7720426	0	5
ins		3,328	.3650841	.4815261	0	1
income		3,328	36.80485	26.70121	-90.05	237.301

1.2. Estimate a LP, Probit and a Logit model to explain dambexp. Store the β coefficients and report them in a table.

```
global xlist age i.female educ i.blhisp totchr i.ins income //Define regressor list $xlist
bys dambexp: su $xlist

** LPM
eststo LPM: reg dambexp $xlist, robust // heterosk needs to be corrected

** probit
eststo probit: probit dambexp $xlist

** logit
eststo logit: logit dambexp $xlist

esttab LPM probit logit, se scalar(N r2 ll) mtitle("LPM" "Probit" "Logit") title(Estimated C
```

-> dambexp = 0

Variable		Obs	Mean	Std. dev.	Min	Max
-----+-----						
age		526	3.695627	1.076467	2.1	6.4
female						
0		526	.7338403	.4423695	0	1
1		526	.2661597	.4423695	0	1
educ		526	12.48859	2.697241	0	17

blhisp						
0		526	.5171103	.5001828	0	1

1		526	.4828897	.5001828	0	1
totchr		526	.0912548	.3074311	0	2
ins						
0		526	.6977186	.4596837	0	1
1		526	.3022814	.4596837	0	1
income		526	31.63409	23.17116	0	166.78

-> dambexp = 1

Variable		Obs	Mean	Std. dev.	Min	Max

age		2,802	4.124697	1.116641	2.1	6.4
female						
0		2,802	.4461099	.4971761	0	1
1		2,802	.5538901	.4971761	0	1
educ		2,802	13.5778	2.513906	0	17
blhisp						
0		2,802	.7241256	.4470336	0	1

1		2,802	.2758744	.4470336	0	1
totchr		2,802	.5567452	.809943	0	5
ins						
0		2,802	.6231263	.4846893	0	1
1		2,802	.3768737	.4846893	0	1
income		2,802	37.77552	27.20742	-90.05	237.301

Linear regression

Number of obs = 3,328
F(7, 3320) = 69.43

Prob > F = 0.0000
 R-squared = 0.1276
 Root MSE = .34114

dambexp	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
age	.0216413	.0056048	3.86	0.000	.0106521	.0326304
1.female	.1394928	.0119061	11.72	0.000	.1161487	.1628368
educ	.0143544	.0025774	5.57	0.000	.009301	.0194078
1.blhisp	-.0889738	.0143088	-6.22	0.000	-.1170288	-.0609188
totchr	.0830088	.0057913	14.33	0.000	.0716539	.0943637
1.ins	.0364663	.0119311	3.06	0.002	.0130733	.0598592
income	.000443	.0002215	2.00	0.046	8.70e-06	.0008774
_cons	.4485313	.0430509	10.42	0.000	.3641224	.5329402

Iteration 0: Log likelihood = -1452.4289
 Iteration 1: Log likelihood = -1218.0426
 Iteration 2: Log likelihood = -1195.6199
 Iteration 3: Log likelihood = -1195.5158
 Iteration 4: Log likelihood = -1195.5158

Probit regression

Number of obs = 3,328
 LR chi2(7) = 513.83
 Prob > chi2 = 0.0000
 Pseudo R2 = 0.1769

Log likelihood = -1195.5158

dambexp	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
age	.0868152	.0274556	3.16	0.002	.0330032	.1406272
1.female	.6635053	.0609648	10.88	0.000	.5440164	.7829941
educ	.061884	.012039	5.14	0.000	.038288	.0854801
1.blhisp	-.3657835	.0619095	-5.91	0.000	-.4871239	-.2444432
totchr	.7957496	.0712174	11.17	0.000	.656166	.9353332
1.ins	.169107	.0629296	2.69	0.007	.0457673	.2924467
income	.0026773	.0013105	2.04	0.041	.0001088	.0052458
_cons	-.6686471	.1941247	-3.44	0.001	-1.049125	-.2881698

Iteration 0: Log likelihood = -1452.4289

Iteration 1: Log likelihood = -1238.914
 Iteration 2: Log likelihood = -1194.7039
 Iteration 3: Log likelihood = -1192.8189
 Iteration 4: Log likelihood = -1192.8089
 Iteration 5: Log likelihood = -1192.8089

Logistic regression

Number of obs = 3,328

LR chi2(7) = 519.24

Prob > chi2 = 0.0000

Log likelihood = -1192.8089

Pseudo R2 = 0.1787

dambexp	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
age	.1618629	.0496641	3.26	0.001	.0645229	.2592028
1.female	1.226724	.1130182	10.85	0.000	1.005212	1.448235
educ	.1080498	.0210874	5.12	0.000	.0667192	.1493804
1.blhisp	-.6666381	.1093362	-6.10	0.000	-.8809332	-.452343
totchr	1.554664	.1499606	10.37	0.000	1.260746	1.848581
1.ins	.296996	.1133154	2.62	0.009	.0749019	.5190901
income	.00462	.0024332	1.90	0.058	-.000149	.009389
_cons	-1.26832	.3407805	-3.72	0.000	-1.936237	-.600402

Estimated Coefficients

	(1) LPM	(2) Probit	(3) Logit
main			
age	0.0216*** (0.00560)	0.0868** (0.0275)	0.162** (0.0497)
0.female	0 (.)	0 (.)	0 (.)
1.female	0.139*** (0.0119)	0.664*** (0.0610)	1.227*** (0.113)
educ	0.0144*** (0.00258)	0.0619*** (0.0120)	0.108*** (0.0211)
0.blhisp	0	0	0

	(.)	(.)	(.)
1.blhisp	-0.0890*** (0.0143)	-0.366*** (0.0619)	-0.667*** (0.109)
totchr	0.0830*** (0.00579)	0.796*** (0.0712)	1.555*** (0.150)
0.ins	0 (.)	0 (.)	0 (.)
1.ins	0.0365** (0.0119)	0.169** (0.0629)	0.297** (0.113)
income	0.000443* (0.000222)	0.00268* (0.00131)	0.00462 (0.00243)
_cons	0.449*** (0.0431)	-0.669*** (0.194)	-1.268*** (0.341)

N	3328	3328	3328
r2	0.128		
ll	-1139.1	-1195.5	-1192.8

Standard errors in parentheses

* p<0.05, ** p<0.01, *** p<0.001

1.3. Estimate the Marginal Effect at the Mean for each model, and report them in a table. You will want to use the `estpost margins` post-estimation command, with the relevant option for MEM. Pay special attention to the treatment of discrete regressors. Hint: check to see any differences in the estimated MEs based on whether you use factor notation; for example, `i.female` vs `female`.

```
est clear

** LPM
qui reg dambexp $xlist, robust
estpost margins, dydx(*) atmean
est store LPM

** probit
qui probit dambexp $xlist
```

```

estpost margins, dydx(*) atmean
est store probit

** logit
qui logit dambexp $xlist
estpost margins, dydx(*) atmean
est store logit

esttab LPM probit logit, se scalar(N r2 ll) mtitle("LPM" "Probit" "Logit") ttitle(Marginal Ef.

```

Conditional marginal effects Number of obs = 3,328
Model VCE: Robust

Expression: Linear prediction, predict()
dy/dx wrt: age 1.female educ 1.blhisp totchr 1.ins income

At: age = 4.056881 (mean)
0.female = .4915865 (mean)
1.female = .5084135 (mean)
educ = 13.40565 (mean)
0.blhisp = .6914063 (mean)
1.blhisp = .3085938 (mean)
totchr = .4831731 (mean)
0.ins = .6349159 (mean)
1.ins = .3650841 (mean)
income = 36.80485 (mean)

		Delta-method					
		dy/dx	std. err.	t	P> t	[95% conf. interval]	
age		.0216413	.0056048	3.86	0.000	.0106521	.0326304
1.female		.1394928	.0119061	11.72	0.000	.1161487	.1628368
educ		.0143544	.0025774	5.57	0.000	.009301	.0194078
1.blhisp		-.0889738	.0143088	-6.22	0.000	-.1170288	-.0609188
totchr		.0830088	.0057913	14.33	0.000	.0716539	.0943637
1.ins		.0364663	.0119311	3.06	0.002	.0130733	.0598592
income		.000443	.0002215	2.00	0.046	8.70e-06	.0008774

Note: dy/dx for factor levels is the discrete change from the base level.

Conditional marginal effects Number of obs = 3,328

Model VCE: OIM

Expression: Pr(dambexp), predict()

dy/dx wrt: age 1.female educ 1.blhisp totchr 1.ins income

At: age = 4.056881 (mean)
0.female = .4915865 (mean)
1.female = .5084135 (mean)
educ = 13.40565 (mean)
0.blhisp = .6914063 (mean)
1.blhisp = .3085938 (mean)
totchr = .4831731 (mean)
0.ins = .6349159 (mean)
1.ins = .3650841 (mean)
income = 36.80485 (mean)

		Delta-method				
		dy/dx	std. err.	z	P> z	[95% conf. interval]

age		.0152201	.004837	3.15	0.002	.0057396 .0247005
1.female		.1184629	.0112862	10.50	0.000	.0963423 .1405835
educ		.0108492	.0021305	5.09	0.000	.0066736 .0150249
1.blhisp		-.0701607	.0130287	-5.39	0.000	-.0956964 -.0446249
totchr		.1395073	.0102098	13.66	0.000	.1194966 .1595181
1.ins		.0288089	.0104412	2.76	0.006	.0083445 .0492733
income		.0004694	.0002295	2.05	0.041	.0000195 .0009192

Note: dy/dx for factor levels is the discrete change from the base level.

Conditional marginal effects

Number of obs = 3,328

Model VCE: OIM

Expression: Pr(dambexp), predict()

dy/dx wrt: age 1.female educ 1.blhisp totchr 1.ins income

At: age = 4.056881 (mean)
0.female = .4915865 (mean)
1.female = .5084135 (mean)
educ = 13.40565 (mean)
0.blhisp = .6914063 (mean)
1.blhisp = .3085938 (mean)
totchr = .4831731 (mean)
0.ins = .6349159 (mean)
1.ins = .3650841 (mean)

income = 36.80485 (mean)

	Delta-method					
	dy/dx	std. err.	z	P> z	[95% conf. interval]	
age	.0135771	.0042044	3.23	0.001	.0053365	.0218176
1.female	.1068865	.0107295	9.96	0.000	.085857	.1279159
educ	.0090632	.0018074	5.01	0.000	.0055208	.0126056
1.blhisp	-.062462	.0116115	-5.38	0.000	-.0852201	-.039704
totchr	.1304052	.0093686	13.92	0.000	.112043	.1487674
1.ins	.0241537	.0089994	2.68	0.007	.0065151	.0417922
income	.0003875	.0002043	1.90	0.058	-.0000128	.0007879

Note: dy/dx for factor levels is the discrete change from the base level.

Marginal Effects at the Mean

	(1) LPM	(2) Probit	(3) Logit
age	0.0216*** (0.00560)	0.0152** (0.00484)	0.0136** (0.00420)
0.female	0 (.)	0 (.)	0 (.)
1.female	0.139*** (0.0119)	0.118*** (0.0113)	0.107*** (0.0107)
educ	0.0144*** (0.00258)	0.0108*** (0.00213)	0.00906*** (0.00181)
0.blhisp	0 (.)	0 (.)	0 (.)
1.blhisp	-0.0890*** (0.0143)	-0.0702*** (0.0130)	-0.0625*** (0.0116)
totchr	0.0830*** (0.00579)	0.140*** (0.0102)	0.130*** (0.00937)
0.ins	0	0	0

	(.)	(.)	(.)
1.ins	0.0365** (0.0119)	0.0288** (0.0104)	0.0242** (0.00900)
income	0.000443* (0.000222)	0.000469* (0.000230)	0.000388 (0.000204)

N	3328	3328	3328
r2	0.128		
ll	-1139.1	-1195.5	-1192.8

Standard errors in parentheses

* p<0.05, ** p<0.01, *** p<0.001

1.4. Estimate the Average Marginal Effect at the Mean for each model, and report them in a table. You will want to use the `estpost margins` post-estimation command, with the relevant option for AME.

```
est clear

** LPM
qui reg dambexp $xlist, robust
estpost margins, dydx(*)
est store LPM

** probit
qui probit dambexp $xlist
estpost margins, dydx(*)
est store probit

** logit
qui logit dambexp $xlist
estpost margins, dydx(*)
est store logit

esttab LPM probit logit, se scalar(N r2 ll) mtitle("LPM" "Probit" "Logit") title(Average Marginal Effects)
```

Average marginal effects
Model VCE: Robust

Number of obs = 3,328

Expression: Linear prediction, predict()
dy/dx wrt: age 1.female educ 1.blhisp totchr 1.ins income

		Delta-method				[95% conf. interval]	
		dy/dx	std. err.	t	P> t		
age		.0216413	.0056048	3.86	0.000	.0106521	.0326304
1.female		.1394928	.0119061	11.72	0.000	.1161487	.1628368
educ		.0143544	.0025774	5.57	0.000	.009301	.0194078
1.blhisp		-.0889738	.0143088	-6.22	0.000	-.1170288	-.0609188
totchr		.0830088	.0057913	14.33	0.000	.0716539	.0943637
1.ins		.0364663	.0119311	3.06	0.002	.0130733	.0598592
income		.000443	.0002215	2.00	0.046	8.70e-06	.0008774

Note: dy/dx for factor levels is the discrete change from the base level.

Average marginal effects
Model VCE: OIM

Number of obs = 3,328

Expression: Pr(dambexp), predict()
dy/dx wrt: age 1.female educ 1.blhisp totchr 1.ins income

		Delta-method				[95% conf. interval]	
		dy/dx	std. err.	z	P> z		
age		.0173895	.0054832	3.17	0.002	.0066426	.0281365
1.female		.132906	.0118133	11.25	0.000	.1097524	.1560596
educ		.0123957	.0023862	5.19	0.000	.0077189	.0170725
1.blhisp		-.0777517	.0137954	-5.64	0.000	-.1047901	-.0507133
totchr		.1593929	.0139062	11.46	0.000	.1321371	.1866486
1.ins		.033324	.0121638	2.74	0.006	.0094834	.0571646
income		.0005363	.0002621	2.05	0.041	.0000225	.0010501

Note: dy/dx for factor levels is the discrete change from the base level.

Average marginal effects
Model VCE: OIM

Number of obs = 3,328

Expression: Pr(dambexp), predict()
dy/dx wrt: age 1.female educ 1.blhisp totchr 1.ins income

	Delta-method					
	dy/dx	std. err.	z	P> z	[95% conf. interval]	
age	.0181007	.0055261	3.28	0.001	.0072697	.0289317
1.female	.1357958	.0117423	11.56	0.000	.1127813	.1588102
educ	.0120829	.0023248	5.20	0.000	.0075264	.0166394
1.blhisp	-.0795676	.0136638	-5.82	0.000	-.1063481	-.0527871
totchr	.1738536	.0164518	10.57	0.000	.1416087	.2060985
1.ins	.0326182	.0121752	2.68	0.007	.0087552	.0564812
income	.0005166	.0002718	1.90	0.057	-.000016	.0010493

Note: dy/dx for factor levels is the discrete change from the base level.

Average Marginal Effects

	(1) LPM	(2) Probit	(3) Logit
age	0.0216*** (0.00560)	0.0174** (0.00548)	0.0181** (0.00553)
0.female	0 (.)	0 (.)	0 (.)
1.female	0.139*** (0.0119)	0.133*** (0.0118)	0.136*** (0.0117)
educ	0.0144*** (0.00258)	0.0124*** (0.00239)	0.0121*** (0.00232)
0.blhisp	0 (.)	0 (.)	0 (.)
1.blhisp	-0.0890*** (0.0143)	-0.0778*** (0.0138)	-0.0796*** (0.0137)
totchr	0.0830*** (0.00579)	0.159*** (0.0139)	0.174*** (0.0165)
0.ins	0 (.)	0 (.)	0 (.)

1.ins	0.0365** (0.0119)	0.0333** (0.0122)	0.0326** (0.0122)
income	0.000443* (0.000222)	0.000536* (0.000262)	0.000517 (0.000272)

N	3328	3328	3328
r2	0.128		
ll	-1139.1	-1195.5	-1192.8

Standard errors in parentheses

* p<0.05, ** p<0.01, *** p<0.001

1.5. Check to see how well the probit model predicts the outcome using the `estat classification` post-estimation command.

```
qui probit dambexp age i.female educ i.blhisp totchr i.ins income
estat classification
```

Probit model for dambexp

		----- True -----		
Classified		D	~D	Total
-----+-----+-----+-----				
+		2768	469	3237
-		34	57	91
-----+-----+-----+-----				
Total		2802	526	3328

Classified + if predicted Pr(D) >= .5

True D defined as dambexp != 0

Sensitivity	Pr(+ D)		98.79%
Specificity	Pr(- ~D)		10.84%
Positive predictive value	Pr(D +)		85.51%
Negative predictive value	Pr(~D -)		62.64%

False + rate for true ~D	Pr(+ ~D)		89.16%
False - rate for true D	Pr(- D)		1.21%
False + rate for classified +	Pr(~D +)		14.49%
False - rate for classified -	Pr(D -)		37.36%

Correctly classified	84.89%
----------------------	--------

1.6. Construct and interpret the LR test for the omission of income in the probit model. Do this in two ways: (1) using the post estimation `lrtest`; (2) manually recreate (1)'s results (both test-statistic and p-value).

```
est clear

** Remove income from xlist
global xlist age i.female educ i.blhisp totchr i.ins

eststo modU: qui probit dambexp $xlist income
scalar logl_U = e(ll)

eststo modR: qui probit dambexp $xlist
scalar logl_R = e(ll)

lrtest modU modR

** Replicate
scalar stat = 2 * (logl_U - logl_R)
scalar pval = chi2tail(1,stat)
scalar list
```

Likelihood-ratio test
Assumption: modR nested within modU

```
LR chi2(1) = 4.30
Prob > chi2 = 0.0382
    pval = .03817363
    stat = 4.297269
    logl_R = -1197.6644
    logl_U = -1195.5158
```

Part 2

Estimate the following models for `lambexp` treating the selection into non-zero `lambexp` value as endogenous using, both Heckman 2-step method and also MLE.

In the main data `lambexp` is missing for values of `ambexp=0`. Before proceeding,

```
replace lambexp = 0 if ambexp==0
```

(526 real changes made)

This will correction will also treat observations with `ambexp=1` as equivalent to `=0`; however, this is only a single observation.

```
** Remove income from xlist
global xlist age i.female educ i.blhisp totchr i.ins
```

2.1. Estimate the Heckman 2-step estimator and store the results. In addition, store the Mills ratio as a separate variable. Use `income` as the excluded variable. This means that `income` appears in the selection equation, but NOT the main equation.

```
eststo heck_2sW: heckman lambexp $xlist, select(dambexp = $xlist income) twostep mills(mills)
```

```
Heckman selection model -- two-step estimates    Number of obs    =      3,328
(regression model with sample selection)         Selected       =      2,802
                                                Nonselected     =       526

                                                Wald chi2(6)     =      193.43
                                                Prob > chi2      =       0.0000
```

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
lambexp						
age	.2024668	.0242202	8.36	0.000	.1549961	.2499374
1.female	.2921341	.0725756	4.03	0.000	.1498886	.4343796
educ	.0123889	.0115682	1.07	0.284	-.0102844	.0350622
1.blhisp	-.1828659	.0653449	-2.80	0.005	-.3109396	-.0547922
totchr	.5006332	.0485548	10.31	0.000	.4054675	.5957988
1.ins	-.0465097	.0529742	-0.88	0.380	-.1503373	.0573179
_cons	5.288927	.288522	18.33	0.000	4.723435	5.85442
dambexp						
age	.0868152	.0274556	3.16	0.002	.0330032	.1406272
1.female	.6635053	.0609648	10.88	0.000	.5440165	.7829941

educ		.061884	.012039	5.14	0.000	.038288	.0854801
1.blhisp		-.3657835	.0619095	-5.91	0.000	-.4871239	-.2444432
totchr		.7957496	.0712174	11.17	0.000	.656166	.9353332
1.ins		.169107	.0629296	2.69	0.007	.0457673	.2924467
income		.0026773	.0013105	2.04	0.041	.0001088	.0052458
_cons		-.6686471	.1941247	-3.44	0.001	-1.049125	-.2881698

/mills							
lambda		-.4637133	.2825997	-1.64	0.101	-1.017598	.090172

rho		-0.35907					
sigma		1.2914258					

2.2. Replicate these results by applying the following steps: (1) estimate the selection equation using a probit model; (2) create the mills ratio; (3) **compare** your mills ratio with the one stored above; (4) estimate the main equation, including the mills ratio.

```
probit dambexp $xlist income
predict index, xb
gen mills = normalden(index)/normal(index)
compare mills mills_a
reg lambexp $xlist mills
```

```
Iteration 0: Log likelihood = -1452.4289
Iteration 1: Log likelihood = -1218.0426
Iteration 2: Log likelihood = -1195.6199
Iteration 3: Log likelihood = -1195.5158
Iteration 4: Log likelihood = -1195.5158
```

Probit regression

Number of obs = 3,328
LR chi2(7) = 513.83
Prob > chi2 = 0.0000
Pseudo R2 = 0.1769

Log likelihood = -1195.5158

dambexp		Coefficient	Std. err.	z	P> z	[95% conf. interval]

age		.0868152	.0274556	3.16	0.002	.0330032 .1406272
1.female		.6635053	.0609648	10.88	0.000	.5440164 .7829941
educ		.061884	.012039	5.14	0.000	.038288 .0854801

1.blhisp		-.3657835	.0619095	-5.91	0.000	-.4871239	-.2444432
totchr		.7957496	.0712174	11.17	0.000	.656166	.9353332
1.ins		.169107	.0629296	2.69	0.007	.0457673	.2924467
income		.0026773	.0013105	2.04	0.041	.0001088	.0052458
_cons		-.6686471	.1941247	-3.44	0.001	-1.049125	-.2881698

		----- Difference -----		
	Count	Minimum	Average	Maximum
mills<mills_a	1660	-5.02e-08	-8.91e-09	-1.75e-14
mills>mills_a	1668	7.75e-15	8.64e-09	5.45e-08
Jointly defined	3328	-5.02e-08	-1.10e-10	5.45e-08
Total	3328			

Source		SS	df	MS	Number of obs	=	3,328
Model		6325.70678	7	903.672398	F(7, 3320)	=	164.14
Residual		18278.1125	3,320	5.50545557	Prob > F	=	0.0000
Total		24603.8193	3,327	7.39519666	R-squared	=	0.2571
					Adj R-squared	=	0.2555
					Root MSE	=	2.3464

lambexp		Coefficient	Std. err.	t	P> t	[95% conf. interval]	
age		.1628716	.0410937	3.96	0.000	.0823001	.2434431
1.female		.1898127	.1257893	1.51	0.131	-.0568198	.4364451
educ		-.0016555	.0199775	-0.08	0.934	-.0408251	.037514
1.blhisp		-.1086689	.1107824	-0.98	0.327	-.3258776	.1085397
totchr		.4202052	.0839337	5.01	0.000	.2556382	.5847723
1.ins		-.0686172	.0899672	-0.76	0.446	-.2450141	.1077796
mills		-4.592193	.4582359	-10.02	0.000	-5.490646	-3.693739
_cons		5.904903	.5004624	11.80	0.000	4.923657	6.886149

2.3 Estimate the marginal effects of the selection equation. You can do this using the `margins` command, with `predict()` option `pse1`. This should correspond to a probit model estimation above.

```
qui heckman lambexp $xlist, select(dambexp = $xlist income) twostep
margins, dydx(*) predict(psel)
```

Average marginal effects
Model VCE: Conventional

Number of obs = 3,328

Expression: Pr(dambexp), predict(psel)
dy/dx wrt: age 1.female educ 1.blhisp totchr 1.ins income

		Delta-method				[95% conf. interval]	
		dy/dx	std. err.	z	P> z		
age		.0173895	.0054832	3.17	0.002	.0066426	.0281365
1.female		.132906	.0118133	11.25	0.000	.1097524	.1560596
educ		.0123957	.0023862	5.19	0.000	.0077189	.0170725
1.blhisp		-.0777517	.0137954	-5.64	0.000	-.1047901	-.0507133
totchr		.1593929	.0139062	11.46	0.000	.1321371	.1866486
1.ins		.033324	.0121638	2.74	0.006	.0094834	.0571646
income		.0005363	.0002621	2.05	0.041	.0000225	.0010501

Note: dy/dx for factor levels is the discrete change from the base level.

2.4. Estimate the Maximum Likelihood version of the Heckmann correction (with an excluded variable) and store the results.

```
eststo heck_mlW: heckman lambexp $xlist, select(dambexp = $xlist income) nolog mills(mills_a
```

Heckman selection model	Number of obs	=	3,328
(regression model with sample selection)	Selected	=	2,802
	Nonselected	=	526
	Wald chi2(6)	=	288.88
Log likelihood = -5836.219	Prob > chi2	=	0.0000

		Coefficient	Std. err.	z	P> z	[95% conf. interval]	
lambexp							

age		.2119749	.0230072	9.21	0.000	.1668816	.2570682
1.female		.3481441	.0601142	5.79	0.000	.2303223	.4659658
educ		.018716	.0105473	1.77	0.076	-.0019563	.0393883
1.blhisp		-.2185714	.0596687	-3.66	0.000	-.3355199	-.101623
totchr		.53992	.0393324	13.73	0.000	.4628299	.61701
1.ins		-.0299871	.0510882	-0.59	0.557	-.1301182	.0701439
_cons		5.044056	.2281259	22.11	0.000	4.596938	5.491175

dambexp							
age		.0879359	.027421	3.21	0.001	.0341917	.14168
1.female		.6626649	.0609384	10.87	0.000	.5432278	.7821021
educ		.0619485	.0120295	5.15	0.000	.0383711	.0855258
1.blhisp		-.3639377	.0618734	-5.88	0.000	-.4852073	-.2426682
totchr		.7969518	.0711306	11.20	0.000	.6575383	.9363653
1.ins		.1701367	.0628711	2.71	0.007	.0469117	.2933618
income		.0027078	.0013168	2.06	0.040	.000127	.0052886
_cons		-.6760546	.1940288	-3.48	0.000	-1.056344	-.2957652

/athrho		-.1313456	.1496292	-0.88	0.380	-.4246134	.1619222
/lnsigma		.2398173	.0144598	16.59	0.000	.2114767	.268158

rho		-.1305955	.1470772			-.4008098	.1605217
sigma		1.271017	.0183786			1.235501	1.307554
lambda		-.1659891	.1878698			-.5342072	.2022291

LR test of indep. eqns. (rho = 0): chi2(1) = 0.91					Prob > chi2 = 0.3406		

2.5 Compute the marginal effects of each regressor for: (1) probability of selection; (2) the expected value of the outcome; and (3) the expected value of the outcome, conditional on selection. You will need to use the post-estimation command `margins, dydx(*) predict()` with predict options: `psel`, `yexpected`, and `ycond`.

```
qui heckman lambexp $xlist, select(dambexp = $xlist income) nolog

margins, dydx(*) predict(psel)
margins, dydx(*) predict(yexpected)
margins, dydx(*) predict(ycond)
```

Average marginal effects
Model VCE: OIM

Number of obs = 3,328

Expression: Pr(dambexp), predict(psel)
dy/dx wrt: age 1.female educ 1.blhisp totchr 1.ins income

		Delta-method				[95% conf. interval]	
		dy/dx	std. err.	z	P> z		
age		.0176149	.0054761	3.22	0.001	.006882	.0283479
1.female		.1327517	.0118078	11.24	0.000	.1096089	.1558945
educ		.0124093	.0023845	5.20	0.000	.0077357	.0170828
1.blhisp		-.0773377	.0137795	-5.61	0.000	-.1043449	-.0503305
totchr		.159642	.013898	11.49	0.000	.1324024	.1868817
1.ins		.033526	.0121515	2.76	0.006	.0097095	.0573425
income		.0005424	.0002634	2.06	0.039	.0000262	.0010586

Note: dy/dx for factor levels is the discrete change from the base level.

Average marginal effects
Model VCE: OIM

Number of obs = 3,328

Expression: E(lambexp*Pr(dambexp)), predict(yexpected)
dy/dx wrt: age 1.female educ 1.blhisp totchr 1.ins income

		Delta-method				[95% conf. interval]	
		dy/dx	std. err.	z	P> z		
age		.2897346	.0381846	7.59	0.000	.2148942	.364575
1.female		1.14081	.0832165	13.71	0.000	.9777089	1.303911
educ		.0941877	.0167238	5.63	0.000	.0614096	.1269658
1.blhisp		-.6692709	.095073	-7.04	0.000	-.8556106	-.4829312
totchr		1.463455	.0890935	16.43	0.000	1.288834	1.638075
1.ins		.1863924	.0856166	2.18	0.029	.0185871	.3541978
income		.0034285	.0016722	2.05	0.040	.0001511	.0067059

Note: dy/dx for factor levels is the discrete change from the base level.

Average marginal effects
Model VCE: OIM

Number of obs = 3,328

Expression: E(lambexp|Zg>0), predict(ycond)
dy/dx wrt: age 1.female educ 1.blhisp totchr 1.ins income

		Delta-method				[95% conf. interval]	
		dy/dx	std. err.	z	P> z		
age		.2165793	.0222037	9.75	0.000	.1730609	.2600977
1.female		.3826391	.0486389	7.87	0.000	.2873087	.4779695
educ		.0219597	.0097532	2.25	0.024	.0028438	.0410755
1.blhisp		-.2385954	.0551014	-4.33	0.000	-.3465921	-.1305986
totchr		.5816493	.0379133	15.34	0.000	.5073406	.6559579
1.ins		-.0212273	.0499484	-0.42	0.671	-.1191243	.0766697
income		.0001418	.0001762	0.80	0.421	-.0002036	.0004872

Note: dy/dx for factor levels is the discrete change from the base level.

2.6. Now re-estimate the two-step and MLE approach without an excluded variable, storing the results each time. This means that the same set of regressors enter both equations. i.e. include income in the outcome equation.

```
global xlist age i.female educ i.blhisp totchr i.ins income
eststo heck_2sW0: heckman lambexp $xlist, select(dambexp = $xlist) twostep mills(mills_b)
eststo heck_mlw0: heckman lambexp $xlist, select(dambexp = $xlist) nolog mills(mills_b_mle)
```

```
Heckman selection model -- two-step estimates    Number of obs    =      3,328
(regression model with sample selection)         Selected       =      2,802
                                                Nonselected     =       526

                                                Wald chi2(7)    =      192.92
                                                Prob > chi2     =      0.0000
```

		Coefficient	Std. err.	z	P> z	[95% conf. interval]	
lambexp							
age		.2043022	.0244086	8.37	0.000	.1564622	.2521422
1.female		.2786877	.0750154	3.72	0.000	.1316602	.4257151
educ		.0141631	.0118462	1.20	0.232	-.0090551	.0373812
1.blhisp		-.1797416	.0656337	-2.74	0.006	-.3083812	-.051102
totchr		.4938391	.049539	9.97	0.000	.3967445	.5909337
1.ins		-.0461181	.053113	-0.87	0.385	-.1502176	.0579815

income		-.0007456	.0010158	-0.73	0.463	-.0027367	.0012454
_cons		5.306311	.2901551	18.29	0.000	4.737617	5.875004

dambexp							
age		.0868152	.0274556	3.16	0.002	.0330032	.1406272
1.female		.6635053	.0609648	10.88	0.000	.5440165	.7829941
educ		.061884	.012039	5.14	0.000	.038288	.0854801
1.blhisp		-.3657835	.0619095	-5.91	0.000	-.4871239	-.2444432
totchr		.7957496	.0712174	11.17	0.000	.656166	.9353332
1.ins		.169107	.0629296	2.69	0.007	.0457673	.2924467
income		.0026773	.0013105	2.04	0.041	.0001088	.0052458
_cons		-.6686471	.1941247	-3.44	0.001	-1.049125	-.2881698

/mills							
lambda		-.5087361	.2894687	-1.76	0.079	-1.076084	.0586121

rho		-0.39250					
sigma		1.2961455					

Heckman selection model	Number of obs	=	3,328
(regression model with sample selection)	Selected	=	2,802
	Nonselected	=	526
	Wald chi2(7)	=	285.98
Log likelihood = -5836.09	Prob > chi2	=	0.0000

		Coefficient	Std. err.	z	P> z	[95% conf. interval]

lambexp						
age		.2137594	.0232969	9.18	0.000	.1680983 .2594205
1.female		.342293	.0615522	5.56	0.000	.2216528 .4629332
educ		.0202746	.0110032	1.84	0.065	-.0012913 .0418406
1.blhisp		-.2185104	.0598099	-3.65	0.000	-.3357357 -.1012852
totchr		.5375964	.0398453	13.49	0.000	.459501 .6156918
1.ins		-.0287728	.0511856	-0.56	0.574	-.1290946 .0715491
income		-.0005026	.000989	-0.51	0.611	-.0024411 .0014359
_cons		5.041712	.229726	21.95	0.000	4.591458 5.491967

dambexp						
age		.0878613	.0274099	3.21	0.001	.034139 .1415837
1.female		.6628035	.060929	10.88	0.000	.5433848 .7822223

educ		.0617998	.0120332	5.14	0.000	.0382152	.0853844
1.blhisp		-.3636885	.0618724	-5.88	0.000	-.4849562	-.2424207
totchr		.7968988	.0711265	11.20	0.000	.6574934	.9363041
1.ins		.1699645	.0628669	2.70	0.007	.0467476	.2931815
income		.0027483	.0013209	2.08	0.037	.0001595	.0053372
_cons		-.675346	.1939739	-3.48	0.000	-1.055528	-.295164

/athrho		-.1419126	.1535634	-0.92	0.355	-.4428913	.1590661
/lnsigma		.240186	.0146925	16.35	0.000	.2113892	.2689828

rho		-.1409675	.1505118			-.4160382	.157738
sigma		1.271486	.0186813			1.235393	1.308633
lambda		-.1792382	.1924853			-.5565025	.1980261

LR test of indep. eqns. (rho = 0): chi2(1) = 1.02					Prob > chi2 = 0.3122		

2.7. Create a table that reports the four models alongside one another and compare the results.

```
esttab heck_2sW heck_2sWO heck_mlW heck_mlWO, se scalar(N) mtitle("2-step,w/" "2-step,w/o" "1
```

Heckman Selection Models

	(1)	(2)	(3)	(4)
	2-step,w/	2-step,w/o	ML,w/	ML,w/o

lambexp				
age	0.202*** (0.0242)	0.204*** (0.0244)	0.212*** (0.0230)	0.214*** (0.0233)
1.female	0.292*** (0.0726)	0.279*** (0.0750)	0.348*** (0.0601)	0.342*** (0.0616)
educ	0.0124 (0.0116)	0.0142 (0.0118)	0.0187 (0.0105)	0.0203 (0.0110)
1.blhisp	-0.183** (0.0653)	-0.180** (0.0656)	-0.219*** (0.0597)	-0.219*** (0.0598)
totchr	0.501*** (0.0486)	0.494*** (0.0495)	0.540*** (0.0393)	0.538*** (0.0398)

1.ins	-0.0465 (0.0530)	-0.0461 (0.0531)	-0.0300 (0.0511)	-0.0288 (0.0512)
income		-0.000746 (0.00102)		-0.000503 (0.000989)
_cons	5.289*** (0.289)	5.306*** (0.290)	5.044*** (0.228)	5.042*** (0.230)

dambexp age	0.0868** (0.0275)	0.0868** (0.0275)	0.0879** (0.0274)	0.0879** (0.0274)
1.female	0.664*** (0.0610)	0.664*** (0.0610)	0.663*** (0.0609)	0.663*** (0.0609)
educ	0.0619*** (0.0120)	0.0619*** (0.0120)	0.0619*** (0.0120)	0.0618*** (0.0120)
1.blhisp	-0.366*** (0.0619)	-0.366*** (0.0619)	-0.364*** (0.0619)	-0.364*** (0.0619)
totchr	0.796*** (0.0712)	0.796*** (0.0712)	0.797*** (0.0711)	0.797*** (0.0711)
1.ins	0.169** (0.0629)	0.169** (0.0629)	0.170** (0.0629)	0.170** (0.0629)
income	0.00268* (0.00131)	0.00268* (0.00131)	0.00271* (0.00132)	0.00275* (0.00132)
_cons	-0.669*** (0.194)	-0.669*** (0.194)	-0.676*** (0.194)	-0.675*** (0.194)

/mills lambda	-0.464 (0.283)	-0.509 (0.289)		

/				
athrho			-0.131 (0.150)	-0.142 (0.154)

lnsigma			0.240*** (0.0145)	0.240*** (0.0147)
---------	--	--	----------------------	----------------------

N	3328	3328	3328	3328
---	------	------	------	------

Standard errors in parentheses
 * p<0.05, ** p<0.01, *** p<0.001

Postamble

log close

```

      name: <unnamed>
      log: C:\Users\neil_\OneDrive - University of Warwick\Documents\EC910\we
> bsite\warwick-ec910\problem-sets\ps-6\problem-set-6-log.txt
      log type: smcl
      closed on: 19 Nov 2024, 17:41:19

```