MLE: Categorical and Limited Dependent Variables
Unit 4-2: Endogenous Treatment Effects Models

PS2730-2020

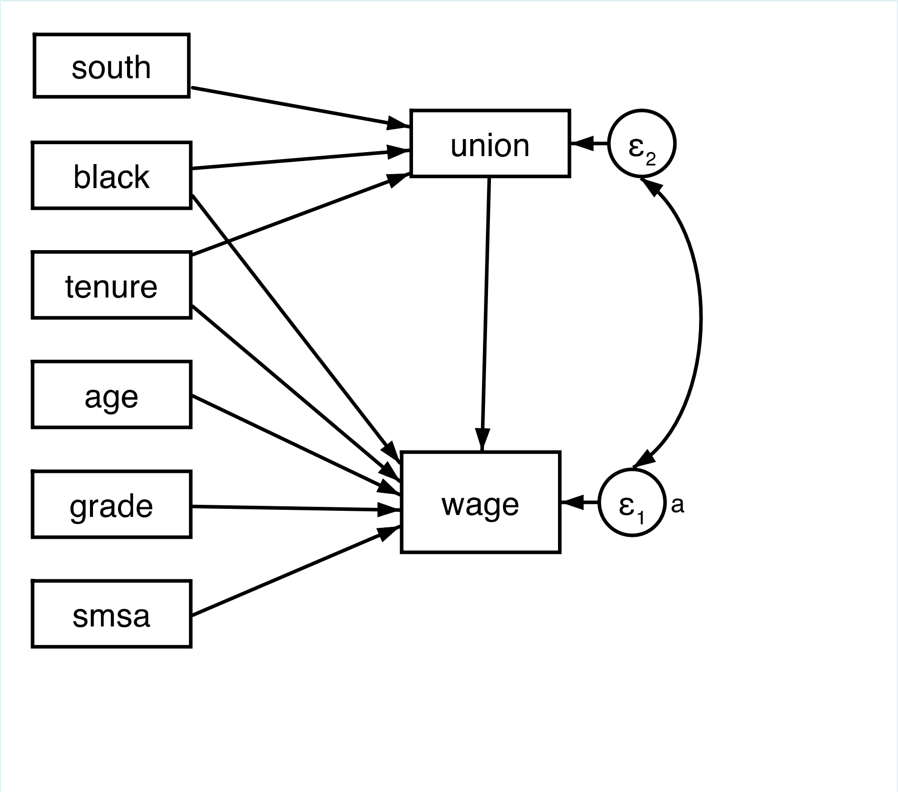
Week 13

Professor Steven Finkel

Endogenous Treatment Effects Model

- Relatively straightforward extension of the sample selection model we just considered. Assume we have dichotomous T representing some kind of treatment, and a continuous outcome Y.
- But the treatment is not randomly assigned; rather units select themselves into the treatment or some other assignment mechanism which *may* result in a treatment-outcome error correlation
- So we have possible endogeneity with T being related to the error term of the outcome equation for Y.
- Formally, the error term of the T equation is related to the error term of the Y equation, exactly as in the figure in week 12, slide 7
- So the estimation of the effect of T is biased unless corrected

- Example: What is the effect of union membership on wages?
- Union membership is likely endogenous to wages – workers select into union jobs based on unobservables which might affect their wages, beyond being in the union



Here is the causal diagram for the two-equation system; some exogenous factors common to both equations but at least *one* exogenous factor ("south" in this case) that causes union membership but not wages to serve as an "instrument" or in a control function regression; this should have a theoretical and defensible basis (in this case, maybe not!)

Estimation of Treatment Effects Model

- Situation is similar to Heckman sample selection model, but here we observe the outcome for all observations, not just those units who were present in the selected sample
- Suggests that we can use a similar two-step **control function** procedure as in the sample selection, using probit to produce an estimate of the error term of the treatment equation, which we can then introduce as a control variable in the outcome equation
- Model

$$\text{Outcome} \quad \mathrm{Y} = XB + \delta \mathrm{T} + \varepsilon$$

$$\text{Treatment} \quad \mathrm{T}^* = \mathrm{W}\gamma + u$$

$$\mathrm{t=1 \ if} \ (\mathrm{W}\gamma + u) > 0$$

$$\mathrm{t=0 \ if} \ (\mathrm{W}\gamma + u) \leq 0$$

- Step 1: Estimate the treatment equation via probit
- Step 2: Generate the "generalized probit residuals" which, assuming a normal distribution for *u*, is represented as λ:

$$\text{For t=1 (treatment): } \lambda = \frac{\phi(W\hat{\gamma})}{\Phi(W\hat{\gamma})}$$

$$\text{For t=0 (control): } \lambda = \frac{\phi(-W\hat{\gamma})}{1-\Phi(W\hat{\gamma})}$$

- λ gives the instantaneous probability of *not* being treated for the treatment group, and of *being* treated for the control group
- As P(t=1) increases for the treatment group, λ *decreases*, and as P(t=0) increases for the treatment group, λ *increases* (to represent the large error term that was needed to push the case over the threshold to be 1 on the treatment variable)
- Opposite for control group: λ *increases* as P(t=1) increases; *decreases* as P(t=0) increases

- Step 3: Estimate the outcome equation with λ as an additional control

  Outcome $Y_i = X_i B + \delta T_i + \rho^* \lambda_i + \varepsilon_i$

- The coefficient for λ, ϱ*, is the estimate of ϱ, the correlation between treatment and outcome error terms, multiplied by the standard deviation of the outcome error ε.

- Can see the impact of ϱ by working out the E(Y|T) equations

$$E(Y|T=1) = X_i\hat{\beta} + \delta + \rho\sigma_\varepsilon \frac{\phi(W_i\gamma)}{\Phi(W_i\gamma)}$$

$$E(Y|T=0) = X_i\hat{\beta} \qquad + \rho\sigma_\varepsilon \frac{\phi(-W_i\gamma)}{1-\Phi(W_i\gamma)}$$

Naive Estimation of Treatment Effect (ΔE(Y|T=1,0))

$$= \delta + \rho\sigma \frac{\phi(W_i\gamma)}{(1-\Phi(W_i\gamma))^2}$$

- As ϱ >0, naïve (OLS) will *overestimate* the true treatment effect (δ); as ϱ <0, naïve (OLS) will *underestimate* the true treatment effect (δ); when ϱ =0 there is no endogeneity in the estimation of the effect of T

```
. etregress wage tenure age black smsa, treat (union=black south tenure)  twostep

Linear regression with endogenous treatment     Number of obs     =       1210
Estimator: two-step                             Wald chi2(7)      =     277.28
                                                Prob > chi2       =     0.0000
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
wage						
tenure	.0359637	.0616359	0.58	0.560	-.0848404	.1567678
age	.1999924	.0212775	9.40	0.000	.1582892	.2416956
black	-1.323392	.2102795	-6.29	0.000	-1.735532	-.9112518
smsa	1.021352	.142392	7.17	0.000	.7422687	1.300435
union	5.889273	1.189625	4.95	0.000	3.557652	8.220895
_cons	-.7269618	.5230607	-1.39	0.165	-1.752142	.2982185
union						
black	.4397974	.0972261	4.52	0.000	.2492377	.6303572
south	-.4895032	.0933276	-5.24	0.000	-.6724221	-.3065844
tenure	.0997638	.0236575	4.22	0.000	.053396	.1461317
_cons	-.9679795	.0746464	-12.97	0.000	-1.114284	-.8216753
hazard						
lambda	-2.88192	.6841896	-4.21	0.000	-4.222907	-1.540933
rho	-1.00000					
sigma	2.8234253					

ML Estimation of Treatment Effects Model

- Maddala (1983) and others argue that Heckman two-step is less efficient than ML estimation of the model. Also Heckman is highly sensitive to violations of normality assumptions and model misspecification in the first stage, so sensitivity analysis is normally recommended to test robustness of the results
- ML estimation: two models, one for treated units, one for control

$$\text{Outcome} \quad \mathrm{Y} = XB + \delta \mathrm{T} + \varepsilon$$

$$\text{Treatment} \quad \mathrm{T}^{*} = \mathrm{W}\gamma + u$$

$$\text{t=1 if } (\mathrm{W}\gamma + u) > 0$$

$$\text{t=0 if } (\mathrm{W}\gamma + u) \leq 0$$

$$\text{Treated (t=1)} \qquad \mathrm{Y} = XB + (\mathrm{W}\gamma + u)\delta + \varepsilon$$

$$\text{Untreated (t=0)} \qquad \mathrm{Y} = XB + \varepsilon$$

- So likelihood function looks like this:

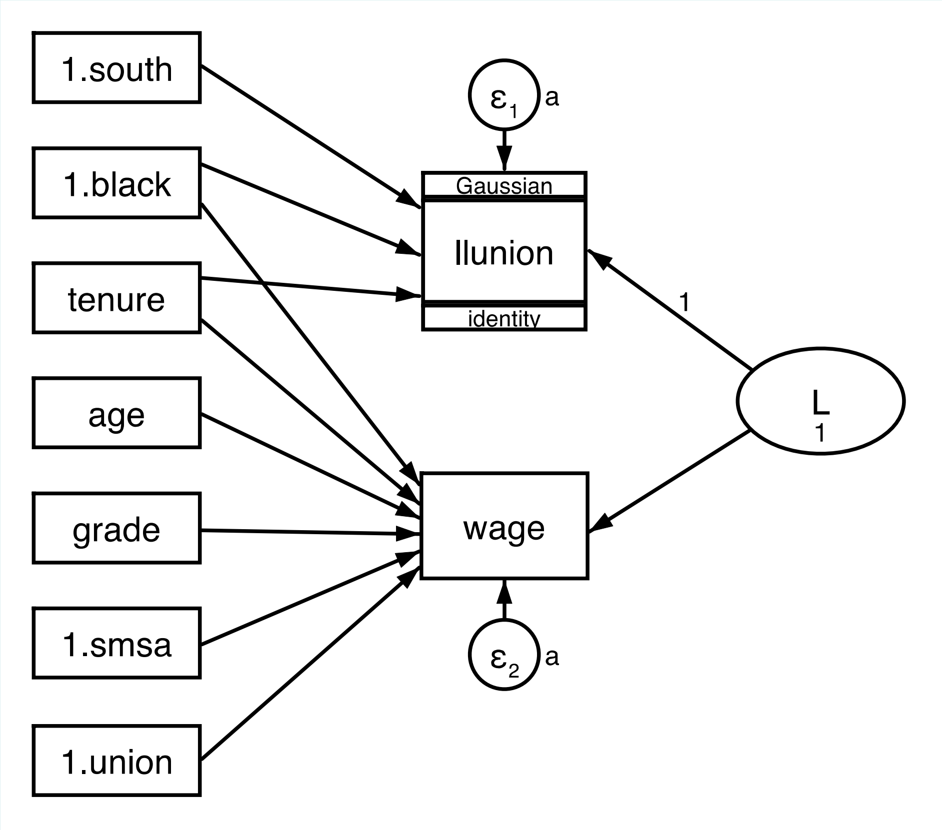
Treated (t=1)
$$\ln\Phi\left(\frac{-w_i\gamma+(y_i-x_i\beta-\delta)\rho\sigma_\varepsilon}{\sqrt{1-\rho^2}}\right)-\frac{1}{2}\frac{(y_i-x_i\beta-\delta)^2}{\sigma_\varepsilon}-\ln\sqrt{2\pi\sigma_\varepsilon}$$

Untreated (t=0)
$$\ln\Phi\left(\frac{-w_i\gamma(y_i-x_i\beta)\rho\sigma_\varepsilon}{\sqrt{1-\rho^2}}\right)-\frac{1}{2}\frac{(y_i-x_i\beta-\delta)^2}{\sigma_\varepsilon}-\ln\sqrt{2\pi\sigma_\varepsilon}$$

- Sum up for treated and untreated, maximize wrt to $\beta,\gamma,\varrho,\sigma$!

ML Estimation of Treatment Effects Model

- In Stata's Generalized Structural Equation Model (GSEM): Add an unobserved latent variable L to represent ϱ, the outcome-treatment error term correlation. Need to "trick" Stata into allowing an error term for the treatment equation's probit, which normally has no error (see do file)



```
Log likelihood =  -3051.575

 ( 1)  [llunion]L = 1
 ( 2)  - [/]var(e.wage) + [/]var(e.llunion) = 0
 ( 3)  [/]var(L) = 1
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
llunion						
1.south	-.8542673	.136439	-6.26	0.000	-1.121683	-.5868518
1.black	.6704049	.148057	4.53	0.000	.3802185	.9605913
tenure	.1282024	.0357986	3.58	0.000	.0580384	.1983664
L	1	(constrained)				
_cons	-1.302676	.1407538	-9.25	0.000	-1.578548	-1.026804
wage						
1.black	-.7882472	.1367077	-5.77	0.000	-1.056189	-.520305
tenure	.1524015	.0369595	4.12	0.000	.0799621	.2248408
age	.1487409	.0193291	7.70	0.000	.1108566	.1866252
grade	.4205658	.0293577	14.33	0.000	.3630258	.4781057
1.smsa	.9117044	.1249041	7.30	0.000	.6668969	1.156512
1.union	2.945816	.2749549	10.71	0.000	2.406914	3.484718
L	-1.706795	.1288024	-13.25	0.000	-1.959243	-1.454347
_cons	-4.351572	.5283952	-8.24	0.000	-5.387207	-3.315936
var(L)	1	(constrained)				
var(e.wage)	1.163821	.2433321			.7725324	1.753298
var(e.llunion)	1.163821	.2433321			.7725324	1.753298

Critical coefficients:

1) Effect of L (the latent variable) on wage: -1.70 significant at .05 level -- that is the equivalent of the rho correlation between errors (actually 1.16/-1.70=-.68)

2) Effect of endogenous union treatment on wave: 2.95, much lower than two-step estimate

- Alternative estimation: Stata "Extended Regression" module (ERM)
- Handles all sorts of outcome variables (continuous, dichotomous, ordered, censored) with endogenous covariates and endogenous treatments of various kinds (continuous, dichotomous, ordered), endogenous sample selection
- Also handles panel data which includes random effects for units or other multilevel structuring in the data
- Estimates from ML (default) or option for Heckman two-step
- Our example:

  etregress wage tenure age black smsa, treat (union=black south tenure)

```
Linear regression with endogenous treatment       Number of obs     =      1,210
Estimator: maximum likelihood                     Wald chi2(5)      =     447.17
Log likelihood = -3146.6831                       Prob > chi2       =     0.0000
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
wage						
tenure	.1167843	.040619	2.88	0.004	.0371725	.196396
age	.1908645	.0206146	9.26	0.000	.1504606	.2312684
black	-1.116333	.1492317	-7.48	0.000	-1.408821	-.8238439
smsa	1.088846	.1345847	8.09	0.000	.8250646	1.352627
1.union	3.388961	.2609577	12.99	0.000	2.877494	3.900429
_cons	-.2546132	.4801038	-0.53	0.596	-1.195599	.6863729
union						
black	.4692957	.0953045	4.92	0.000	.2825024	.656089
south	-.6209058	.082777	-7.50	0.000	-.7831457	-.4586659
tenure	.0824715	.0230314	3.58	0.000	.0373309	.1276122
_cons	-.8556358	.0716176	-11.95	0.000	-.9960038	-.7152677
/athrho	-.7646964	.0815895	-9.37	0.000	-.9246088	-.604784
/lnsigma	.8077796	.0285642	28.28	0.000	.7517948	.8637644
rho	-.6438349	.0477687			-.7280704	-.540445
sigma	2.242922	.0640673			2.120803	2.372073
lambda	-1.444072	.1392013			-1.716901	-1.171242

```
LR test of indep. eqns. (rho = 0):   chi2(1) =    32.60   Prob > chi2 = 0.0000
```

- Crucial coefficients

1) rho=-.64 with small standard error
2) LR test of the independence of the outcome and treatment equations is a test that rho=0; here rejected
3) Effect of union on wage=3.39, close to GSEM estimate

Extensions: Binary Outcome

- Can extend these models to apply to different combinations of continuous/dichotomous endogenous regressors and outcomes
- IV regression or 2SRI (control function regression) for continuous/continuous
- ML or Heckman treatment effects for dichotomous/continuous
- For continuous/dichotomous, two-step control function or ML as options, with additional possibility of using "special regressors" that satisfy assumptions of IV regression
- These "special regressor" models use heteroskedasticity or other distributional properties of the endogenous regressor to identify the model
- See Lewbel (2000; 2012) for more on this work

- Two-Step method: continuous endogenous regressor (X) and dichotomous outcome (D)

$$\text{Outcome} \qquad D^* = XB + \beta X^e + \varepsilon$$

$$\text{Endogenous} \quad X^e = W\gamma + u$$

With W and X having some but not all elements in common, i.e., X^e must be a function of at least one variable that is not a cause of the outcome D*

- Then, assuming ε and *u* are jointly normal:

$$\text{Endogenous} \quad \hat{X}^e = W\gamma$$

$$\hat{u} = X^e - \hat{X}^e$$

$$\text{Outcome} \qquad D^* = XB + \beta X^e + \lambda\hat{u} + \varepsilon$$

- This is a straightforward probit regression in the second stage outcome; the *u* controls for the endogenous portion of X^e and so β represents the "true" effect of X^e

$$D^* = XB + \beta X^e + \lambda \hat{u} + \varepsilon$$

$$D = 1 \text{ if } XB + \beta X^e + \lambda \hat{u} + \varepsilon > 0$$

$$D = 0 \text{ if } XB + \beta X^e + \lambda \hat{u} + \varepsilon \leq 0$$

$$P(D = 1) = P(\varepsilon > -(XB + \beta X^e + \lambda \hat{u})$$

$$P(D = 1) = \Phi(XB + \beta X^e + \lambda \hat{u})$$

- This is the model implemented in Stata "ivprobit", two-step
- ML version as "ivprobit" without two-step, or in ERM as "eprobit"

```
. ivprobit union black south (tenure=age smsa), twostep
Checking reduced-form model...

Two-step probit with endogenous regressors        Number of obs   =      1,210
                                                  Wald chi2(3)    =      40.09
                                                  Prob > chi2     =     0.0000
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
tenure	.111052	.080076	1.39	0.165	-.0458941	.2679982
black	.4394138	.0972647	4.52	0.000	.2487785	.6300491
south	-.4872009	.0945017	-5.16	0.000	-.6724209	-.3019809
_cons	-.9909486	.1730404	-5.73	0.000	-1.330102	-.6517956

```
Instrumented:  tenure
Instruments:   black south age smsa

Wald test of exogeneity: chi2(1) = 0.02                   Prob > chi2 = 0.8824
```

← Two-step

ML – Extended Regression →

```
. eprobit union black south , endog(tenure=age smsa)

Iteration 0:   log likelihood = -2877.9554
Iteration 1:   log likelihood = -2877.9552

Extended probit regression                        Number of obs   =      1,210
                                                  Wald chi2(3)    =      38.60
Log likelihood = -2877.9552                       Prob > chi2     =     0.0000
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
union						
black	.4392008	.0972898	4.51	0.000	.2485164	.6298852
south	-.4898916	.0933104	-5.25	0.000	-.6727766	-.3070066
tenure	.1127219	.0804917	1.40	0.161	-.045039	.2704828
_cons	-.9929996	.1652216	-6.01	0.000	-1.316828	-.6691711
tenure						
age	.1752129	.0157299	11.14	0.000	.1443829	.2060429
smsa	.1617086	.1063658	1.52	0.128	-.0467645	.3701817
_cons	-2.234716	.3769622	-5.93	0.000	-2.973548	-1.495883
var(e.tenure)	2.560817	.1041119			2.36468	2.773223
corr(e.tenure,e.union)	-.0229516	.1365242	-0.17	0.866	-.2827599	.2399942

Note: no endogeneity in this model, according to insignificant error correlation in eprobit, and insignificant Wald test in ivprobit

Extension: Binary Treatment and Binary Outcome

- Endgenous treatment and binary (dichotomous outcome) normally estimated via ML, not control function methods

Outcome $D^* = XB + \delta T + \varepsilon$

Endogenous $T^* = W\gamma + u$

$D=1$ if $XB + \delta T > \varepsilon$

$T=1$ if $W\gamma > u$

$(\varepsilon, u) \sim N(0, \Sigma)$

- Where Σ is the covariance between the errors of the outcome and treatment equations (containing rho(ϱ) in previous slides)
- As in all of these models: need to have some elements in W that are not in X – these serve the same function as instrumental variables in previous models

- This model implemented as "biprobit" in Stata:

biprobit (union =black south tenure collgrad) (collgrad=age sms)

```
Seemingly unrelated bivariate probit            Number of obs     =      1,210
                                                Wald chi2(6)      =     109.65
Log likelihood = -964.15959                     Prob > chi2       =     0.0000
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
union						
black	.4504212	.0987426	4.56	0.000	.2568892	.6439532
south	-.4835541	.0932949	-5.18	0.000	-.6664088	-.3006994
tenure	.0974279	.0268456	3.63	0.000	.0448115	.1500443
collgrad	.5600467	.9367598	0.60	0.550	-1.275969	2.396062
_cons	-1.023733	.0899562	-11.38	0.000	-1.200043	-.8474216
collgrad						
age	.1353075	.0195513	6.92	0.000	.0969877	.1736273
smsa	.2312438	.1251775	1.85	0.065	-.0140996	.4765873
_cons	-4.696822	.4944085	-9.50	0.000	-5.665845	-3.7278
/athrho	-.1856135	.5186246	-0.36	0.720	-1.202099	.830872
rho	-.1835109	.5011592			-.8342937	.6809439

```
LR test of rho=0: chi2(1) = .157773                     Prob > chi2 = 0.6912
```

Correlation between "collgrad" and "union" equations=-.18, not significant, and collgrad has no significant effect on union membership

- More options available within ERM. Endogenous treatment in "eprobit" allows different error correlations between treatment and outcome equations for treatment and control groups, for example; also allows estimation of potential outcomes and treatment effects via different outcome equations for treatment and control groups

eprobit union black south tenure , entreat (collgrad=c.age c.sms, nointer pocorr) vce(robust)

```
Extended probit regression                      Number of obs     =      1,210
                                                Wald chi2(4)      =      63.06
Log pseudolikelihood = -963.56978               Prob > chi2       =     0.0000
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
union						
black	.4462026	.0981253	4.55	0.000	.2538806	.6385246
south	-.4813505	.0925093	-5.20	0.000	-.6626655	-.3000356
tenure	.0992165	.025419	3.90	0.000	.0493962	.1490368
1.collgrad	1.255879	.6841022	1.84	0.066	-.0849366	2.596695
_cons	-.9826396	.0998021	-9.85	0.000	-1.178248	-.787031
collgrad						
age	.1338979	.0149357	8.96	0.000	.1046245	.1631713
smsa	.2354379	.1219552	1.93	0.054	-.0035899	.4744658
_cons	-4.664532	.3814617	-12.23	0.000	-5.412183	-3.91688
corr(e.collgrad,e.union)						
0.collgrad	.0661744	.3942153	0.17	0.867	-.6105357	.6870346
1.collgrad	-.5897724	.3407418	-1.73	0.083	-.9355775	.3334593

Error correlation between college and union equations significant (p<.10) only for college group; effect of college graduate now significant also (p<.10)

- Extended Regression (ERM) has additional capabilities:
  - Ordinal outcomes
  - Endogeneous ordinal and censored covariates and treatment effects
  - All of the above with additional controls for sample selection