Computer Lab in Economics
Master in International Economics
Advanced Topics with Stata

Inmaculada Álvarez Ayuso

Office 314 (Módulo I) www.uam.es/inmaculada.alvarez

E-mail: inmaculada.alvarez@uam.es

Discrete choice models

When the dependent variable is qualitative, we can use several cross section methods:

Logit Regression (Logit, logistic)

Probabilistic Regression (Probit, ivprobit)

Nested models (nlogit)

And the extensión with panel data:

xtlogit and xtprobit

Discrete Choice Models (Logistic models, logit, logistic)

In Stata we estimate the following expression:

$$\Pr(y_j \neq 0 \mid \mathbf{x}_j) = \frac{\exp(\mathbf{x}_j\beta)}{1+\exp(\mathbf{x}_j\beta)}$$

Where 0 indicates the negative outcome and the rest of values the positive outcome.

Discrete Choice Models (Logistic models)

asclogit	asclogit	Alternative-specific conditional logit model
blogit	glogit	Logit regression for grouped data
clogit	clogit	Conditional (fixed-effects) logistic regression
exlogistic	exlogistic	Exact logistic regression
glogit	glogit	Weighted least-squares logistic regression for grouped data
logit	logit	Logistic regression, reporting coefficients
melogit	melogit	Multilevel mixed-effects logistic regression
mlogit	mlogit	Multinomial (polytomous) logistic regression
nlogit	nlogit	Nested logit regression (RUM-consistent and nonnormalized)
ologit	ologit	Ordered logistic regression
rologit	rologit	Rank-ordered logistic regression
scobit	scobit	Skewed logistic regression
slogit	slogit	Stereotype logistic regression

Discrete Choice Models (Probabilistic models, probit)

The dependent variable values that are not zeros are interpreted as sucess. Therefore, the model probit estimate the probability of positive outcome estimating in Stata:

$$\Pr(y_j \neq 0 \mid \mathbf{x}_j) = \Phi(\mathbf{x}_j\beta)$$

Where Φ is the standard cumulative normal

Discrete choice models and censored regression models with Stata

Discrete Choice Models (Probabilistic models, probit)

asmprobit	asmprobit	Alternative-specific multinomial probit regression
asroprobit	asroprobit	Alternative-specific rank-ordered probit regression
biprobit	biprobit	Bivariate probit regression
bprobit	glogit	Probit regression for grouped data
gprobit	glogit	Weighted least-squares probit regression for grouped data
heckoprobit	heckoprobit	Ordered probit model with sample selection
heckprobit	heckprobit	Probit model with sample selection
hetprobit	hetprobit	Heteroskedastic probit model
ivprobit	ivprobit	Probit model with endogenous regressors
meprobit	meprobit	Multilevel mixed-effects probit regression
mprobit	mprobit	Multinomial probit regression
oprobit	oprobit	Ordered probit regression
probit	probit	Probit regression

Discrete choice models and censored regression models with Stata

Discrete Choice Models (Panel Data, xt)

xtcloglog	xtcloglog	Random-effects and population-averaged cloglog models
xtgee	xtgee	GEE population-averaged generalized linear models
xtlogit	xtlogit	Fixed-effects, random-effects, and population-averaged logit models
xtologit	xtologit	Random-effects ordered logistic models
xtoprobit	xtoprobit	Random-effects ordered probit models
xtprobit	xtprobit	Random-effects and population-averaged probit models

Discrete choice models (Logistic Models, logit)

```
. logit d time lnlabour lncapital

Iteration 0:   log likelihood = -381.92319
Iteration 1:   log likelihood = -93.589244
Iteration 2:   log likelihood = -56.127125
Iteration 3:   log likelihood = -39.327086
Iteration 4:   log likelihood = -38.687417
Iteration 5:   log likelihood = -38.685318
Iteration 6:   log likelihood = -38.685317

Logistic regression                             Number of obs     =        551
                                                LR chi2(3)        =     686.48
                                                Prob > chi2       =     0.0000
Log likelihood = -38.685317                     Pseudo R2         =     0.8987
```

dlngdp	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
time	.1875598	.0975078	1.92	0.054	-.003552	.3786716
lnlabour	5.734696	2.753282	2.08	0.037	.3383618	11.13103
lncapital	12.25291	3.501039	3.50	0.000	5.391005	19.11482
_cons	-279.285	49.73825	-5.62	0.000	-376.7702	-181.7998

```
Note: 100 failures and 103 successes completely determined.
```

Discrete choice models (Logistic models, logistic)

```
. logistic d time lnlabour lncapital

Logistic regression                                     Number of obs   =        551
                                                        LR chi2(3)      =     686.48
                                                        Prob > chi2     =     0.0000
Log likelihood = -38.685317                             Pseudo R2       =     0.8987

------------------------------------------------------------------------------
       dlngdp | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
         time |   1.206302   .1176239     1.92   0.054     .9964543    1.460343
    lnlabour |   309.4189   851.9176     2.08   0.037     1.402648    68256.66
   lncapital |   209591.2   733786.7     3.50   0.000     219.4237    2.00e+08
       _cons |   5.1e-122   2.5e-120    -5.62   0.000     2.3e-164    1.11e-79
------------------------------------------------------------------------------
Note: 100 failures and 103 successes completely determined.
```

Discrete choice models (Probit models, probit)

```
. probit d time lnlabour lncapital

Iteration 0:   log likelihood = -381.92319
Iteration 1:   log likelihood =  -91.88237
Iteration 2:   log likelihood = -51.296539
Iteration 3:   log likelihood = -38.772294
Iteration 4:   log likelihood =  -38.36075
Iteration 5:   log likelihood = -38.357171
Iteration 6:   log likelihood =  -38.35717

Probit regression                                 Number of obs   =        551
                                                  LR chi2(3)      =     687.13
                                                  Prob > chi2     =     0.0000
Log likelihood =  -38.35717                       Pseudo R2       =     0.8996
```

dlngdp	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
time	.1123379	.0542427	2.07	0.038	.0060241	.2186517
lnlabour	3.365344	1.546807	2.18	0.030	.3336575	6.397031
lncapital	6.687844	1.851378	3.61	0.000	3.05921	10.31648
_cons	-155.6562	25.50266	-6.10	0.000	-205.6405	-105.6719

```
Note: 159 failures and 169 successes completely determined.
```

Discrete choice models (IV Probit models, ivprobit)

ivprobit d time lnlabour (lncapital=L.lncapital)

Probit model with endogenous regressors — Number of obs = 532
Wald chi2(3) = 43.92
Log likelihood = 1382.6461 — Prob > chi2 = 0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lncapital	7.117596	1.896232	3.75	0.000	3.401049	10.83414
time	.1033933	.0540672	1.91	0.056	-.0025765	.2093632
lnlabour	2.915511	1.64579	1.77	0.076	-.3101788	6.141201
_cons	-156.7908	25.62216	-6.12	0.000	-207.0094	-106.5723
/athrho	-.1922423	.1680494	-1.14	0.253	-.5216131	.1371284
/lnsigma	-4.088596	.030657	-133.37	0.000	-4.148683	-4.028509
rho	-.1899086	.1619887			-.4789441	.1362753
sigma	.0167628	.0005139			.0157852	.0178008

Instrumented: lncapital
Instruments: time lnlabour L.lncapital

Wald test of exogeneity (/athrho = 0): chi2(1) = 1.31 Prob > chi2 = 0.2526

note: 63 failures and 20 successes completely determined.

Discrete choice models
(IV Probit models, ivprobit)

```
. ivprobit d time lnlabour (lncapital=L.lncapital), twostep
Checking reduced-form model...

Two-step probit with endogenous regressors        Number of obs   =        532
                                                  Wald chi2(3)    =      40.99
                                                  Prob > chi2     =     0.0000
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lncapital	7.249522	1.973768	3.67	0.000	3.381008	11.11804
time	.1053097	.0549411	1.92	0.055	-.0023729	.2129924
lnlabour	2.969551	1.668286	1.78	0.075	-.3002299	6.239331
_cons	-159.697	26.87606	-5.94	0.000	-212.3731	-107.0209

```
Instrumented:  lncapital
Instruments:   time lnlabour L.lncapital

Wald test of exogeneity:     chi2(1) =     1.28          Prob > chi2 = 0.2582

note: 64 failures and 23 successes completely determined.
```

Discrete choice models (panel data, xtlogit)

```
. xtlogit dtime time lnlabour lncapital,fe
note: multiple positive outcomes within groups encountered.
note: 15 groups (435 obs) dropped because of all positive or
      all negative outcomes.

Iteration 0:   log likelihood = -38.119269
Iteration 1:   log likelihood =  -34.74298
Iteration 2:   log likelihood = -34.422131
Iteration 3:   log likelihood =  -34.42158
Iteration 4:   log likelihood =  -34.42158

Conditional fixed-effects logistic regression   Number of obs      =       116
Group variable: region                          Number of groups   =         4

                                                Obs per group: min =        29
                                                               avg =      29.0
                                                               max =        29

                                                LR chi2(3)        =     12.71
Log likelihood  =  -34.42158                    Prob > chi2       =    0.0053
```

dtime	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
time	-.2862795	.1616971	-1.77	0.077	-.6032	.0306409
lnlabour	15.57461	6.462829	2.41	0.016	2.907697	28.24152
lncapital	8.241422	4.738497	1.74	0.082	-1.045862	17.52871

Discrete choice models (panel data, xtlogit)

Random-effects logistic regression — Number of obs = 551
Group variable: **region** — Number of groups = 19

Random effects u_i ~ **Gaussian** — Obs per group: min = 29
avg = 29.0
max = 29

Wald chi2(3) = 12.63
Log likelihood = -47.574502 — Prob > chi2 = 0.0055

dtime	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
time	-.2428463	.1494689	-1.62	0.104	-.5358	.0501073
lnlabour	15.25416	5.223641	2.92	0.003	5.016013	25.49231
lncapital	6.847959	4.357607	1.57	0.116	-1.692793	15.38871
_cons	-307.9764	90.34114	-3.41	0.001	-485.0418	-130.911
/lnsig2u	1.349541	.8666249			-.349013	3.048094
sigma_u	1.963582	.8508445			.8398714	4.590767
rho	.5395902	.2152979			.1765556	.8649757

Likelihood-ratio test of rho=0: chibar2(01) = 23.52 Prob >= chibar2 = 0.000

Discrete choice models (panel data, xtlogit)

```
. xtlogit dtime time lnlabour lncapital,pa

Iteration 1: tolerance = 1.7910262
Iteration 2: tolerance = .03578409
Iteration 3: tolerance = .02635525
Iteration 4: tolerance = .00993613
Iteration 5: tolerance = .00294689
Iteration 6: tolerance = .00105796
Iteration 7: tolerance = .00027303
Iteration 8: tolerance = .00003712
Iteration 9: tolerance = .0000223
Iteration 10: tolerance = .00001198
Iteration 11: tolerance = 5.788e-06
Iteration 12: tolerance = 2.078e-06
Iteration 13: tolerance = 5.677e-07

GEE population-averaged model                   Number of obs      =       551
Group variable:                      region     Number of groups   =        19
Link:                                 logit     Obs per group: min =        29
Family:                            binomial                    avg =      29.0
Correlation:                   exchangeable                    max =        29
                                                Wald chi2(3)       =     13.23
Scale parameter:                          1     Prob > chi2        =    0.0042
```

dtime	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
time	-.0941489	.0714552	-1.32	0.188	-.2341986	.0459008
lnlabour	8.133919	3.048706	2.67	0.008	2.158564	14.10927
lncapital	2.436501	2.128691	1.14	0.252	-1.735656	6.608657
_cons	-144.841	40.17136	-3.61	0.000	-223.5755	-66.10662

Discrete choice models (panel data, xtprobit)

Random-effects probit regression — Number of obs = 551
Group variable: **region** — Number of groups = 19

Random effects u_i ~ **Gaussian** — Obs per group: min = 29
avg = 29.0
max = 29

Wald chi2(3) = 12.23
Log likelihood = -47.312859 — Prob > chi2 = 0.0066

dtime	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
time	-.1339346	.0849038	-1.58	0.115	-.300343	.0324738
lnlabour	8.730887	2.939871	2.97	0.003	2.968847	14.49293
lncapital	3.747883	2.479123	1.51	0.131	-1.111109	8.606874
_cons	-173.4767	52.28395	-3.32	0.001	-275.9514	-71.00208
/lnsig2u	.1852775	.8599265			-1.500148	1.870702
sigma_u	1.097065	.4716978			.4723317	2.548108
rho	.5461873	.2131472			.1824035	.8665395

Likelihood-ratio test of rho=0: chibar2(01) = 23.81 Prob >= chibar2 = 0.000

Discrete choice models (panel data, xtprobit)

```
. xtprobit dtime time lnlabour lncapital, pa

Iteration 1: tolerance = 1.0537426
Iteration 2: tolerance = .10779412
Iteration 3: tolerance = .04057797
Iteration 4: tolerance = .01527505
Iteration 5: tolerance = .00639187
Iteration 6: tolerance = .00291959
Iteration 7: tolerance = .00135327
Iteration 8: tolerance = .0006013
Iteration 9: tolerance = .00024693
Iteration 10: tolerance = .00009044
Iteration 11: tolerance = .00003121
Iteration 12: tolerance = 9.671e-06
Iteration 13: tolerance = 2.024e-06
Iteration 14: tolerance = 1.599e-06
Iteration 15: tolerance = 1.152e-06
Iteration 16: tolerance = 6.267e-07

GEE population-averaged model                   Number of obs      =       551
Group variable:                      region     Number of groups   =        19
Link:                                probit     Obs per group: min =        29
Family:                            binomial                    avg =      29.0
Correlation:                   exchangeable                    max =        29
                                                Wald chi2(3)       =     16.93
Scale parameter:                          1     Prob > chi2        =    0.0007
```

dtime	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
time	-.0560494	.0395893	-1.42	0.157	-.1336429	.0215442
lnlabour	4.973142	1.6627	2.99	0.003	1.71431	8.231975
lncapital	1.466665	1.164106	1.26	0.208	-.814941	3.74827
_cons	-88.23135	21.58646	-4.09	0.000	-130.54	-45.92267

Censored regression models(Tobit models, Tobit)

Title

[R] tobit — Tobit regression

Syntax

tobit *depvar* [*indepvars*] [*if*] [*in*] [*weight*] , **ll**[(#)] **ul**[(#)] [*options*]

options	description
Model	
noconstant	suppress constant term
* **ll**[(#)]	left-censoring limit
* **ul**[(#)]	right-censoring limit
offset(*varname*)	include *varname* in model with coefficient constrained to 1
SE/Robust	
vce(*vcetype*)	*vcetype* may be **oim**, **robust**, **cluster** *clustvar*, **bootstrap**, or **jackknife**
Reporting	
level(#)	set confidence level; default is **level**(95)
display_options	control spacing and display of omitted variables and base and empty cells
Maximization	
maximize_options	control the maximization process; seldom used
+ **coeflegend**	display coefficients' legend instead of coefficient table

Censored regression models(Tobit models, Tobit)

Title

[R] tobit postestimation — Postestimation tools for tobit

Description

The following postestimation commands are available for **tobit**:

	command	description
	estat	AIC, BIC, VCE, and estimation sample summary
	estat (svy)	postestimation statistics for survey data
	estimates	cataloging estimation results
	hausman	Hausman's specification test
	lincom	point estimates, standard errors, testing, and inference for linear combinations of coefficients
	linktest	link test for model specification
(1)	lrtest	likelihood-ratio test
	margins	marginal means, predictive margins, marginal effects, and average marginal effects
	nlcom	point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients
	predict	predictions, residuals, influence statistics, and other diagnostic measures
	predictnl	point estimates, standard errors, testing, and inference for generalized predictions
	suest	seemingly unrelated estimation
	test	Wald tests of simple and composite linear hypotheses
	testnl	Wald tests of nonlinear hypotheses

(1) **lrtest** is not appropriate with **svy** estimation results.

Censored regression models (Panel dataTobit models, xttobit)

Syntax

xttobit *depvar* [*indepvars*] [*if*] [*in*] [*weight*] [, *options*]

options	description
Model	
noconstant	suppress constant term
ll(*varname*|*#***)**	left-censoring variable/limit
ul(*varname*|*#***)**	right-censoring variable/limit
offset(*varname***)**	include *varname* in model with coefficient constrained to 1
constraints(*constraints***)**	apply specified linear constraints
collinear	keep collinear variables
SE	
vce(*vcetype***)**	*vcetype* may be **oim**, **bootstrap**, or **jackknife**
Reporting	
level(*#***)**	set confidence level; default is **level(95)**
tobit	perform likelihood-ratio test comparing against pooled tobit model
noskip	perform likelihood-ratio test
nocnsreport	do not display constraints
display_options	control spacing and display of omitted variables and base and empty cells
Integration	
intmethod(*intmethod***)**	integration method; *intmethod* may be **mvaghermite**, **aghermite**, or **ghermite**; default is **intmethod(mvaghermite)**
intpoints(*#***)**	use # quadrature points; default is **intpoints(12)**
Maximization	
maximize_options	control the maximization process; seldom used
+ **coeflegend**	display coefficients' legend instead of coefficient table

Censored regression models (Panel dataTobit models. xttobit)

Title

[XT] xttobit postestimation — Postestimation tools for xttobit

Description

The following postestimation commands are available for **xttobit**:

command	description
estat	AIC, BIC, VCE, and estimation sample summary
estimates	cataloging estimation results
lincom	point estimates, standard errors, testing, and inference for linear combinations of coefficients
lrtest	likelihood-ratio test
margins	marginal means, predictive margins, marginal effects, and average marginal effects
nlcom	point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients
predict	predictions, residuals, influence statistics, and other diagnostic measures
predictnl	point estimates, standard errors, testing, and inference for generalized predictions
test	Wald tests of simple and composite linear hypotheses
testnl	Wald tests of nonlinear hypotheses

Simultaneous equation models (SEM), reg3

Title

[R] **reg3** — Three-stage estimation for systems of simultaneous equations

Syntax

Basic syntax

```
reg3 (depvar1 varlist1) (depvar2 varlist2) ...(depvarN varlistN) [if]
     [in] [weight]
```

Full syntax

```
reg3 ([eqname1:]depvar1a [depvar1b ...=]varlist1 [, noconstant])
     ([eqname2:]depvar2a [depvar2b ...=]varlist2 [, noconstant])
     ...
     ([eqnameN:]depvarNa [depvarNb ...=]varlistN [, noconstant])
     [if] [in] [weight] [, options]
```

reg3 allows the estimation of a SEM model with constraints and several estimation methods:

- 2sls and 3sls (by default) for two or three steps
- ols, Ordinary least Squeares
- sure, Semi unrelated equations

Simultaneous equation models (SEM), reg3

Prieto, A., Zofío, J.L. and Alvarez, I. (2014): "Cost economies, urban patterns and population density: The case of public infrastructure for basic utilities". Papers in Regional Science, forthcoming. doi:10.1111/pirs.12096

$$\ln C = \alpha_0 + \sum_{g=1}^{Y} \alpha_g \ln Y_g + \sum_{i=1}^{P} \beta_i \ln P_i + \sum_{k=1}^{Z} \delta_k \ln Z_k$$
$$+ \frac{1}{2}\left[\sum_{g=1}^{Y}\sum_{h=1}^{Y} \alpha_{gh} \ln Y_g \ln Y_h + \sum_{i=1}^{P}\sum_{j=1}^{P} \beta_{ij} \ln P_i \ln P_j + \sum_{k=1}^{Z}\sum_{l=1}^{Z} \delta_{kl} \ln Z_k \ln Z_l \right]$$
$$+ \sum_{g=1}^{Y}\sum_{i=1}^{P} \varphi_{gi} \ln Y_g \ln P_i + \sum_{g=1}^{Y}\sum_{k=1}^{Z} \theta_{gk} \ln Y_g \ln Z_k + \sum_{i=1}^{P}\sum_{k=1}^{Z} \omega_{ik} \ln P_i \ln Z_k.$$

$$S_i = \beta_i + \sum_{j=1}^{P} \beta_{ij} \ln P_j + \sum_{g=1}^{G} \varphi_{gi} \ln Y_g + \sum_{k=1}^{Z} \omega_{ik} \ln Z_k, \quad i = 1, \ldots, P.$$

Simultaneous equation models (SEM), reg3

Table 2. Cost determinants of urban infrastructure (parameter estimates)

Variables	Parameters	S1. Water supply		S2. Sewerage and cleansing of waters		S3. Paving and lighting	
		Coefficients	T-Stat.	Coefficients	T-Stat.	Coefficients	T-Stat.
Constant	α_0	0.078	4.450	0.085	6.540	0.157	6.260
ln Y_1	α_1	0.213	7.050	0.374	11.290	0.107	1.750
ln Y_2	α_2	0.390	14.120	0.119	4.080	0.081	1.660
ln P_1	β_1	0.116	34.210	0.475	73.280	0.927	45.140
ln P_2	β_2	0.703	153.960	0.108	31.250	0.073	35.670
ln P_3	β_3	0.182	55.630	0.416	73.320	–	–
ln Z_1	δ_1	0.152	7.830	0.417	17.520	0.447	9.460
ln Z_2	δ_2	0.201	9.110	0.062	5.160	0.236	9.750
$(\ln Y_1)^2$	α_{11}	−0.091	−0.960	0.363	3.070	−0.147	−0.770
$(\ln Y_2)^2$	α_{22}	0.100	1.500	0.018	0.210	−0.192	−2.120
ln Y_1 ln Y_2	α_{12}	−0.114	−0.820	−0.282	−1.830	0.264	1.290
$(\ln P_1)^2$	β_{11}	0.013	0.920	0.270	6.190	−0.003	−0.600
$(\ln P_2)^2$	β_{22}	−0.181	−4.180	0.002	0.080	−0.003	−0.600
$(\ln P_3)^2$	β_{33}	−0.123	−5.170	0.173	6.250	–	–
ln P_1 ln P_2	β_{12}	0.022	1.140	−0.049	−1.950	0.003	0.600
ln P_1 ln P_3	β_{13}	−0.035	−2.790	−0.220	−7.210	–	–
ln Y_2 ln Z_1	θ_{21}	−0.037	−0.920	0.108	1.430	0.013	0.160
ln Y_2 ln Z_2	θ_{22}	−0.035	−0.810	−0.042	−1.090	0.126	2.870
ln P_1 ln Z_1	ω_{11}	−0.003	−0.560	0.121	7.330	0.032	6.490
ln P_2 ln Z_1	ω_{21}	−0.004	−0.470	−0.006	−0.680	−0.032	−6.540
ln P_3 ln Z_1	ω_{31}	0.007	1.230	−0.115	−7.990	–	–
ln P_1 ln Z_2	ω_{12}	0.048	8.420	−0.064	−7.230	0.016	6.580
ln P_2 ln Z_2	ω_{22}	−0.031	−4.080	0.045	9.510	−0.016	−6.630
ln P_3 ln Z_2	ω_{32}	−0.016	−2.950	0.018	2.400	–	–
F-test		8.77*E+11		4,358.4		16,794.5	
R^2		0.854		0.933		0.709	
Observations		1,793		1,139		1,311	

Source: Own elaboration