

Cross Section Analysis

Dr. Keshab Bhattarai
Business School, University of Hull
Cottingham Road, HU6 7RX
United Kingdom
Email:- K.R.Bhattarai@hull.ac.uk

ADVANCED ECONOMIC THEORY 2007 LECTURE 10

April 16, 2007

Probability Model (Limited Dependent Variables)

- Probit
- Logit
- Multinomial logit
- Tobit
- Sample selection and two stage estimation
- Disequilibrium analysis

Cross Section Analysis: Probability models with Dummy dependent variables

There are many economic applications where a dependent variable takes value of 1 if the event occurs or the answer is yes and 0 if the event does not occur or the answer is no. A few examples as given below can illustrate these issues:

- the labour force participation (1 if a person participates in the labour force, 0 otherwise)
- yes or no vote on a particular issue
- to marry or not to marry
- to study further or to start a job
- to buy or not to buy a particular stock, or commodity
- choice of transportation mode to work (1 if a person drives to work, 0 otherwise)
- Union membership (1 if one is a member of the union, 0 otherwise)
- Owning a house (1 if one owns 0 otherwise)
- Multinomial choices: work as a teacher, or as a clerk, or as a self employed or professional or as a factory worker
- Multinomial ordered choices: strongly agree, agree, neutral, disagree, strongly disagree

Alternative names of such variables in the literature include dichotomous dependent variables, discrete dependent random variable, binary variable, either or choice variable. Algebraically a simplest form of this is given by a linear model as:

$$Y_i = \begin{cases} 1 & \text{if event occurs} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

There are mainly three types of limited dependent variable models

- Linear probability model
- Probit model
- Logit model or multinomial logit or nested logit

There are other type of variables where value of a certain variable is available only for certain observations but not available for others; distributions are truncated or censored. Tobit model handles this issue of selection bias in such models. These models are discussed briefly in section IV.

I. Linear probability model

This is the simplest modelling of limited dependent variable and is given by following function

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad (2)$$

here the dependent variable Y_i takes value of 1 if the event occurs 0 otherwise [e.g. $Y_i = 1$ if a person owns a house, $Y_i = 0$ otherwise; independent variables X_i can be both continuous or dichotomous variables (e.g. family income; or male / female)]

$$E[(Y_i = 1)/X_i] \text{ probability that the event } y \text{ will occur given } X_i$$

$$E[(Y_i = 1)/X_i] = 0(1 - P_i) + 1(P_i) = P_i \quad (3)$$

$$E[(Y_i = 1)/X_i] = \beta_1 + \beta_2 X_i = P_i$$

In theory probability should remain between 0 and 1; where $0 \leq P_i \leq 1$; so the estimated linear probability model should be; $0 \leq E[(Y_i = 1)/X_i] \leq 1$ but this is not always guaranteed in linear model like this.

Numerical example

```

/*=====
Greene Example 19.1. Probability Models
*/=====

```

```

Read ; Nobs = 32 ; nvar = 5 ;
Names = 1 $

```

OBS	GPA	TUCE	PSI	GRADE	32	2.39	19	1	1
1	2.66	20	0	0					
2	2.89	22	0	0					
3	3.28	24	0	0					
4	2.92	12	0	0					
5	4.00	21	0	1					
6	2.86	17	0	0					
7	2.76	17	0	0					
8	2.87	21	0	0					
9	3.03	25	0	0					
10	3.92	29	0	1					
11	2.63	20	0	0					
12	3.32	23	0	0					
13	3.57	23	0	0					
14	3.26	25	0	1					
15	3.53	26	0	0					
16	2.74	19	0	0					
17	2.75	25	0	0					
18	2.83	19	0	0					
19	3.12	23	1	0					
20	3.16	25	1	1					
21	2.06	22	1	0					
22	3.62	28	1	1					
23	2.89	14	1	0					
24	3.51	26	1	0					
25	3.54	24	1	1					
26	2.83	27	1	1					
27	3.39	17	1	1					
28	2.67	24	1	0					
29	3.65	21	1	1					
30	4.00	23	1	1					
31	3.10	21	1	0					

Does attendance to PSI courses improve grade of a student how has attended this?
 Is there any difference in Grade of those who have taken PSI and those who have not taken it?

Result from the estimation as given in the LIMDEP output in the next page show that there were significant effect of PSI attendance; It improves grade by 37 percent.

```
--> Read ; Nobs = 32 ; nvar = 5 ; Names = 1 $
--> Namelist ; X = One,GPA,TUCE,PSI $
--> Matrix ; xbar = mean(x) $
--> Regress ; Lhs = Grade ; Rhs = X $
```

```
-----+-----
| Ordinary least squares regression      Weighting variable = none
| Dep. var. = GRADE      Mean=      .3437500000      , S.D.=      .4825587044
| Model size: Observations =      32, Parameters =      4, Deg.Fr.=      28
| Residuals: Sum of squares= 4.216473951      , Std.Dev.=      .38806
| Fit: R-squared= .415900, Adjusted R-squared =      .35332
| Model test: F[ 3,      28] =      6.65,      Prob value =      .00157
| Diagnostic: Log-L =      -12.9782, Restricted(b=0) Log-L =      -21.5812
|              LogAmemiyaPrCrt.=      -1.775, Akaike Info. Crt.=      1.061
| Autocorrel: Durbin-Watson Statistic =      2.34645,      Rho =      -.17322
|-----+-----
```

```
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|Variable | Coefficient | Standard Error |t-ratio |P[|T|>t] | Mean of X|
|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|Constant | -1.498017120 | .52388862 | -2.859 | .0079 |
|GPA      | .4638516793 | .16195635 | 2.864 | .0078 | 3.1171875
|TUCE     | .1049512224E-01 | .19482854E-01 | .539 | .5944 | 21.937500
|PSI      | .3785547879 | .13917274 | 2.720 | .0111 | .43750000
```

```
--> Calc ; List ; Scale = 1.0 $
SCALE = .10000000000000000D+01
--> Matrix ; List ; ME = Scale * b $
```

Matrix ME has 4 rows and 1 columns.

```
1
+-----+
1 | -.1498017D+01
2 | .4638517D+00
3 | .1049512D-01
4 | .3785548D+00
```

Linear probability model has several problems. Error terms are heteroscedastic; OLS estimator can be unbiased but it not efficient. This can be shown as:

$$u_i = 1 - \beta_1 - \beta_2 X_i \text{ with probability } (1 - P_i) \text{ and}$$

$$u_i = -\beta_1 - \beta_2 X_i \text{ with probability } P_i$$

Variance of the error term is not constant

$$\text{var}(u_i) = (-\beta_1 - \beta_2 X_i)^2 (1 - P_i) + (1 - \beta_1 - \beta_2 X_i)^2 P_i$$

$$\text{var}(u_i) = (-\beta_1 - \beta_2 X_i)^2 (1 - \beta_1 - \beta_2 X_i) + (1 - \beta_1 - \beta_2 X_i)^2 (\beta_1 + \beta_2 X_i)$$

$$\text{var}(u_i) = (\beta_1 + \beta_2 X_i)(1 - \beta_1 - \beta_2 X_i) = P_i(1 - P_i) \quad (4)$$

Thus the variance depends on X_i .

It is possible to transform this model to make it homoscedastic by dividing the original variables by $\sqrt{(\beta_1 + \beta_2 X_i)(1 - \beta_1 - \beta_2 X_i)} = \sqrt{P_i(1 - P_i)} = \sqrt{w_i}$

$$\frac{Y_i}{\sqrt{w_i}} = \frac{\beta_1}{\sqrt{w_i}} + \beta_2 \frac{X_i}{\sqrt{w_i}} + \frac{u_i}{\sqrt{w_i}} \quad (5)$$

Limitations of a linear probability model

- It does not guarantee that the probability lies inside (0,1) bands.
- Probability in non-linear phenomenon: at very low level of income a family does not own a house; at very high level of income **every** one owns a house; marginal effect of income is very negligible. The linear probability model does not explain this fact very well; it violates the fundamental property of probability $0 \leq P_i \leq 1$, as it is possible to have negative as well as greater than 1 value of P_i with a linear model.

II. Probit model

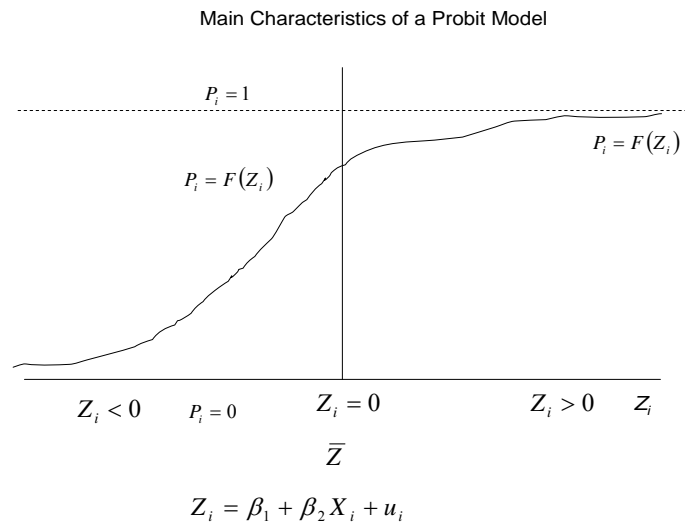
Probit model guarantees that the limiting case of a probability remains between 0 and 1; $0 \leq P_i \leq 1$; making sure that lowest value is not less than zero and the highest value is not more than 1. It is given by following distribution

$$\Pr(Y_i = 1) = \Pr(Z_i^* \leq Z_i) = F(Z_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z_i} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1 + \beta_2 X} e^{-\frac{t^2}{2}} dt \quad (6)$$

Here t is standardised normal variable, $t \sim N(0,1)$. The Unknown parameters of this model β and σ are estimated by the maximum likelihood estimator.

Steps for a probit model

1. Probability depends upon unobserved utility index Z_i which depends upon observable variables such as income and other personal characteristics. There is a threshold of this index after which this affects on occurrence of an event, family starts owning a house, $Z_i \geq Z_i^*$.



Likelihood functions for modelling probability

$$L = \Pr(Y_1)\Pr(Y_2)\dots\Pr(Y_n)$$

$$L = \prod_{i=1}^{n1} P_i \prod_{i=n1}^N (1 - P_i) \quad L = \prod_{i=1}^{n1} P_i^{y_{i,1}} \prod_{i=n1+1}^N P_i^{1-y_i}$$

Log likelihood functions for modelling probability

$$\text{Log}L = \sum_{i=1}^{n1} \log P_i + \sum_{i=1}^N \log(1 - P_i)$$

First order conditions for ML estimates of β and σ

$$\frac{\partial \text{Log}L}{\partial \alpha} = \sum_{i=1}^{n1} \frac{\partial P_i / \partial \alpha}{P_i} - \sum_{i=1}^{n1} \frac{\partial P_i / \partial \alpha}{(1 - P_i)} = 0; \quad \frac{\partial \text{Log}L}{\partial \beta} = \sum_{i=1}^{n1} \frac{\partial P_i / \partial \beta}{P_i} - \sum_{i=1}^{n1} \frac{\partial P_i / \partial \beta}{(1 - P_i)} = 0$$

ML takes the OLS estimates as starting value for computations of optimal β and σ .

$1 - L_0 / L_{\max}$ gives the indication of goodness of fit of ML estimates.

$$L(\alpha, \beta, \sigma) = f(y_1, y_1, \dots, y_n) = \prod_{i=1}^n \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{1}{2}} \exp \left[-\frac{1}{2\sigma^2} (y_i - \alpha - \beta x_i)^2 \right] \quad (1)$$

Take log of this function to get a log-likelihood function.

$$\text{Log}(\alpha, \beta, \sigma) = \left[\sum_{i=1}^n -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (y_i - \alpha - \beta x_i)^2 \right] = c - \frac{n}{2} \log(\sigma^2) - \frac{Q}{2\sigma^2} \quad (2)$$

where $c = -\frac{n}{2} \log(2\pi)$ and $Q = \sum (y_i - \alpha - \beta x_i)^2$

Maximising (2) w.r.t. α , β , σ is equivalent to $Q = \sum (y_i - \alpha - \beta x_i)^2$

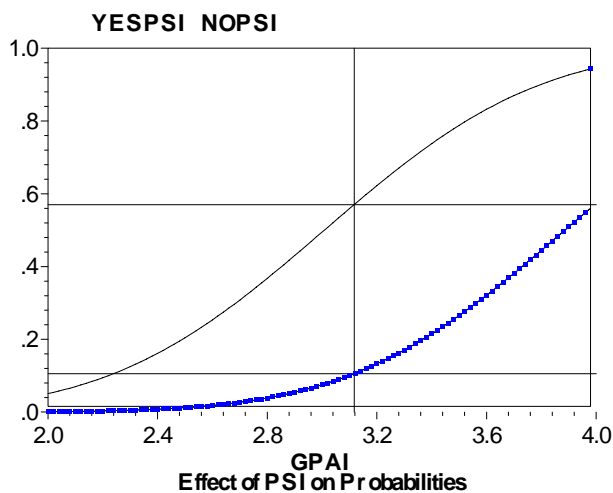
minimizing which is the negative term in the likelihood function. Therefore the estimators of α and β under the ML method are the same as in the OLS method.

If we substitute the values of $\hat{\alpha}$ and $\hat{\beta}$ in (3) the likelihood function just becomes a function of σ

This can be written as

$$\text{Log}(\hat{\alpha}, \hat{\beta}, \sigma) = \sum_{i=1}^n -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2$$

$$\text{Log}(\hat{\alpha}, \hat{\beta}, \sigma) = c - \frac{n}{2} \log(\sigma^2) - \frac{\hat{Q}}{2\sigma^2} = c - \frac{n}{2} \cdot 2 \cdot \log(\sigma) - \frac{\hat{Q}}{2\sigma^2} = c - n \log(\sigma) - \frac{\hat{Q}}{2\sigma^2} \quad (4)$$



Summary on effect of explanatory variable upon probability

The effect of observed variables on probability

$$\frac{\partial P_i}{\partial x_{i,j}} = \begin{cases} \beta_j & \text{for the linear probability model} \\ \beta_j P_j (1 - P_j) & \text{for the logit probability model} \\ \beta_j \phi(Z_i) & \text{for the probit probability model} \end{cases}$$

where $Z_i = \beta_0 + \sum_{i=1}^k \beta_i x_{i,j}$ and $\phi(\cdot)$ is the standard normal density function.

Note: for the logit model:

$$\ln\left(\frac{P_i}{1-P_i}\right) = Z_i = \beta_1 + \beta_2 X_i + u_i \quad \frac{\partial P_i}{\partial X_i} \left[\ln\left(\frac{P_i}{1-P_i}\right) \right] = \beta_2;$$

$$\frac{\partial P_i}{\partial X_i} \left[\ln\left(\frac{P_i}{1-P_i}\right) \right] = \frac{\Delta P_i}{P_i} + \frac{\Delta P_i}{(1-P_i)} = \frac{\Delta P_i}{P_i(1-P_i)} = \beta_2; \quad \Delta P_i = \beta_2 P_i(1-P_i).$$

--> Probit ; Lhs = Grade ; Rhs = X ; Marginal Effects \$

```

-----+-----
Dependent variable is binary, y=0 or y not equal 0
Ordinary least squares regression      Weighting variable = none
Dep. var. = GRADE      Mean=      .3437500000      , S.D.=      .4825587044
Model size: Observations =      32, Parameters =      4, Deg.Fr.=      28
Residuals: Sum of squares= 4.216473951      , Std.Dev.=      .38806
Fit: R-squared= .415900, Adjusted R-squared =      .35332
Model test: F[ 3,      28] =      6.65, Prob value =      .00157
Diagnostic: Log-L =      -12.9782, Restricted(b=0) Log-L =      -21.5812
LogAmemiyaPrCrt.=      -1.775, Akaike Info. Crt.=      1.061
-----+-----

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-1.498017120	.52388862	-2.859	.0042	
GPA	.4638516793	.16195635	2.864	.0042	3.1171875
TUCE	.1049512224E-01	.19482854E-01	.539	.5901	21.937500
PSI	.3785547879	.13917274	2.720	.0065	.43750000

Normal exit from iterations. Exit status=0.

```

-----+-----
Binomial Probit Model
Maximum Likelihood Estimates
Dependent variable          GRADE
Weighting variable          ONE
Number of observations      32
Iterations completed        6
Log likelihood function     -12.81880
Restricted log likelihood   -20.59173
Chi-squared                 15.54585
Degrees of freedom          3
Significance level          .1404896E-02
-----+-----

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Index function for probability					
Constant	-7.452319647	2.5424723	-2.931	.0034	
GPA	1.625810039	.69388249	2.343	.0191	3.1171875
TUCE	.5172894549E-01	.83890261E-01	.617	.5375	21.937500
PSI	1.426332342	.59503790	2.397	.0165	.43750000

```

-----+-----
Partial derivatives of E[y] = F[*] with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Observations used for means are All Obs.
-----+-----

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Index function for probability					
Constant	-2.444733653	.75885194	-3.222	.0013	
GPA	.5333470255	.23246407	2.294	.0218	3.1171875
TUCE	.1696968191E-01	.27119788E-01	.626	.5315	21.937500
PSI	.4679083617	.18764238	2.494	.0126	.43750000

Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability.

Actual	Predicted		Total
	0	1	
0	18	3	21
1	3	8	11
Total	21	11	32

```

--> Calc ; List ; Scale = N01(b'xbar) $
SCALE = .32805002591068580D+00
--> Matrix ; List ; ME = Scale * b $

```

Matrix ME has 4 rows and 1 columns.

```

1 |-----+
2 | -.2444734D+01
3 | .5333470D+00
4 | .1696968D-01
5 | .4679084D+00

```

III. Logit model

Similar to Probit model dependent variable Y_i takes value 1 ($Y_i = 1$) if the event occurs, zero it does not occur. For instance, a student gets a first class mark, value 0 ($Y_i = 0$) otherwise. Probability of getting a first class mark in an exam is a function of

student effort index denoted by Z_i . $P_i = \frac{1}{1 + e^{-Z_i}}$; where $Z_i = \beta_1 + \beta_2 X_i + u_i$ and

$$1 - P_i = 1 - \frac{1}{1 + e^{-Z_i}} = \frac{1 + e^{-Z_i} - 1}{1 + e^{-Z_i}} = \frac{e^{-Z_i}}{1 + e^{-Z_i}} = \frac{1}{1 + e^{Z_i}}; [e = \exp = 2.718].$$

Ratio of odds: $\left[\frac{P_i}{1 - P_i} = \frac{1 + e^{Z_i}}{1 + e^{-Z_i}} = \frac{e^{Z_i}(1 + e^{Z_i})}{e^{Z_i}(1 + e^{-Z_i})} = \frac{e^{Z_i}(1 + e^{Z_i})}{e^{Z_i}(e^{Z_i} + 1)} = e^{Z_i} \right]$; taking log of the

$$\text{odds } \ln\left(\frac{P_i}{1 - P_i}\right) = Z_i$$

Features of a logit model

- Probability goes from 0 to 1 as the index variable Z_i goes from $-\infty$ to $+\infty$.
Probability lies between 0 and 1.
- Log of the odds is linear in x, characteristic variables but probabilities themselves are not linear but non linear function of the parameters. Probabilities are estimated using the maximum likelihood method.
- Any explanatory variable that determines the value of Z_i , measures how the log of odds of an event (i.e. probability of owning a house rather than not owning it) changes as a result of change in explanatory variable such as income.
- We can calculate P_i for given estimates of $\hat{\beta}_1$ and $\hat{\beta}_2$ or all other $\hat{\beta}_i$.
- Limiting case when $P_i=1$; $\ln\left(\frac{P_i}{1 - P_i}\right)$ or when $P_i=0$ $\ln\left(\frac{0}{1 - 1}\right)$; OLS cannot be applied in such case but the maximum likelihood method may be used to estimate the parameters.

An example of a logit model can be illustrated in terms of what determines a first class grade by a student in a given degree such as:

$$Z_i = \beta_1 + \beta_2 H + \beta_3 E + \beta_4 A + \beta_5 P + e_i$$

H = hours of study, E= exercises, A = attendance in lectures and classes; P = papers written for assignment.

--> Logit ; Lhs = Grade ; Rhs = X ; marginal Effects \$

```

+-----+
| Multinomial logit model
| There are 2 outcomes for LH variable GRADE
| These are the OLS start values based on the
| binary variables for each outcome Y(i) = j.
| Coefficients for LHS=0 outcome are set to 0.0
+-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Characteristics in numerator of Prob[Y = 1]					
Constant	-1.498017120	.52388862	-2.859	.0042	
GPA	.4638516793	.16195635	2.864	.0042	3.1171875
TUCE	.1049512224E-01	.19482854E-01	.539	.5901	21.937500
PSI	.3785547879	.13917274	2.720	.0065	.43750000

Normal exit from iterations. Exit status=0.

Multinomial Logit Model	
Maximum Likelihood Estimates	
Dependent variable	GRADE
Weighting variable	ONE
Number of observations	32
Iterations completed	6
Log likelihood function	-12.88963
Restricted log likelihood	-20.59173
Chi-squared	15.40419
Degrees of freedom	3
Significance level	.1501878E-02

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Characteristics in numerator of Prob[Y = 1]					
Constant	-13.02134648	4.9313241	-2.641	.0083	
GPA	2.826112525	1.2629411	2.238	.0252	3.1171875
TUCE	.9515765670E-01	.14155420	.672	.5014	21.937500
PSI	2.378687596	1.0645642	2.234	.0255	.43750000

Partial derivatives of probabilities with respect to the vector of characteristics. They are computed at the means of the Xs. Observations used for means are All Obs.

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Marginal effects on Prob[Y = 1]					
Constant	-2.459760743	.81771031	-3.008	.0026	
GPA	.5338588183	.23703797	2.252	.0243	3.1171875
TUCE	.1797548884E-01	.26236909E-01	.685	.4933	21.937500
PSI	.4493392735	.19676264	2.284	.0224	.43750000

Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability.

Actual	Predicted		Total
	0	1	
0	18	3	21
1	3	8	11
Total	21	11	32

```

* shazam instructions for exercise 18.2
*
sample 1 21
read( table18-1.dat) auto bus x y/skiplines=1
logit y x
stop
* shazam instructions for exercise 18.2
*
dimension y 24 x 24 incr 24
sample 1 21
read(table18-1.dat) auto transit
gen1 ihat=b1+b2+b1*x
distrib ihat / pdf=litf
print incr

```

stop

- **Multinomial choice models:**

1. choice of different brands of a particular goods such as cereals
2. different subjects by students in the business school such as economics, marketing, finance, business, accounting, management or in the university such as science, engineering, medicine, mathematics, philosophy, politics or history, arts or like that.
3. ordered probit or ordered logit for choice of bonds such as AAA BBB; orders are used to rank the outcome.
4. survey questions with ranking.

Let $P_{i,j}$ denote the probability of choosing alternative J by individual i,. There are J alternatives and N individuals.

$$\sum_i^J y_{i,j} = \sum_i^J P_{i,j} = 1$$

$$L = \prod_{i=1}^T P_{i,1}^{y_{i,1}} P_{i,2}^{y_{i,2}} P_{i,3}^{y_{i,3}} \dots P_{i,J}^{y_{i,J}}$$

Individuals make a choice to maximise utility; such utilities are specific to individual and unobserved factors $U_{i,j} = \bar{U}_{i,j} + e_{i,j} = X'_{i,j}\beta + e_{i,j}$. Errors $e_{i,j}$ need to have Weibull distribution a multinomial logit.

Probability that the choice 1 es made when J alternatives were available implies

$$P_{i,1} = \Pr[U_{i,1} > U_{i,2} \text{ and } U_{i,1} > U_{i,3} \dots \text{and } U_{i,1} > U_{i,J}] =$$

$$P_{i,1} = \Pr[e_{i,2} < U_{i,1} - U_{i,2} + e_{i,1} \text{ and } e_{i,3} < U_{i,1} - U_{i,3} + e_{i,1} \dots \text{and } e_{i,J} < U_{i,1} - U_{i,J} + e_{i,1}]$$

$$P_{i,1} = \Pr[e_{i,2} < U_{i,1} - U_{i,2} + e_{i,1} \text{ and } e_{i,3} < U_{i,1} - U_{i,3} + e_{i,1} \dots \text{and } e_{i,J} < U_{i,1} - U_{i,J} + e_{i,1}]$$

$$P_{i,1} = \Pr[e_{i,2} - e_{i,1} < U_{i,1} - U_{i,2} \text{ and } e_{i,3} - e_{i,1} < U_{i,1} - U_{i,3} \dots \text{and } e_{i,J} - e_{i,1} < U_{i,1} - U_{i,J}]$$

Mcfadden (1974) has used multinomial logit model to analyse random utility models.

$$\Pr(e_{i,j} < \varepsilon) = \exp(-\exp(-\varepsilon))$$

$$P_{i,j} = \frac{\exp(X'_{i,j}\beta)}{\sum_{j=1}^J \exp(X'_{i,j}\beta)}$$

Choices are independent of irrelevant alternatives (IIA); only odds between two choices need to be compared when one is confronted with choosing one or another.

$$P_{i,j} = \frac{\exp(X'_{i,j}\beta)}{\sum_{j=1}^J \exp(X'_{i,j}\beta)}$$

$$\frac{P_{i,1}}{P_{i,2}} = \frac{\exp(X'_{i,1}\beta) / \sum_{j=1}^J \exp(X'_{i,j}\beta)}{\exp(X'_{i,2}\beta) / \sum_{j=1}^J \exp(X'_{i,j}\beta)} = \frac{\exp(X'_{i,1}\beta)}{\exp(X'_{i,2}\beta)}$$

$$\frac{P_{i,k}}{P_{i,1}} = \frac{\exp(X'_{i,k}\beta_k)}{\exp(X'_{i,1}\beta_1)} = \exp(X'_{i,k}\beta_k - X'_{i,1}\beta_1) \text{ for } k = 2, \dots, J$$

with normalisation $\beta_1 = 0$ $P_{i,1} = \frac{\exp(X'_{i,j}\beta_1)}{\sum_{j=1}^J \exp(X'_{i,j}\beta_j)} = \frac{1}{1 + \sum_{j=2}^J \exp(X'_{i,j}\beta_j)}$

$$P_{i,1} = \frac{\exp(X'_{i,j}\beta_j)}{\sum_{j=1}^J \exp(X'_{i,j}\beta_j)} = \frac{\exp(X'_{i,j}\beta_j)}{1 + \sum_{j=2}^J \exp(X'_{i,j}\beta_j)} \text{ or } j = 2, \dots, J$$

Parameters in a multinomial logit models are estimated by a maximum likelihood method (see McFadden (1974). Judge, Griffiths, Hill, Lutkepohl and Lee(1990). Similarly one can define the nested logit models.

If there are J alternatives and k subgroups a multiple logit evaluation of probability of choosing j in k subgroup is given by:

$$P_{j,k} = P_j / P_k = \left(\frac{e^{\beta'X|_k}}{\sum_{j=1}^J e^{\beta'X|_k}} \right) \left(\frac{e^{\gamma'z|_k}}{\sum_{k=1}^K e^{\gamma'z|_k}} \right) \left(\frac{\left(\sum_{j=1}^J e^{\beta'X|_k} \right) \sum_{k=1}^K e^{\gamma'z|_k}}{\left(\sum_{k=1}^K \sum_{j=1}^J e^{\beta'X|_k + \gamma'z|_k} \right)} \right)$$

This can be estimated using a two step Maximum likelihood procedure. It can be applied to model the choices of travel modes in which an individual chooses first whether to fly or take a ground transport. Once a ground transport is chosen then similar choice is made for train, bus or car transport. Model can be estimated with limited information maximum likelihood (LIML) or full information maximum likelihood (FIML). (see Greene: 19)

- **Count data models: How many trips to hospital by a patient?**

Poisson random variable $P(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!}$

where λ denotes the intensity of occurrence or the rate parameter and $y = 0,1,2,3,\dots$, denote the counts of events in a given time interval. Mean and variance are the same $E(Y) = \lambda$ and variance $Var(Y) = \lambda$.

$$\lambda_i = \exp(X_i' \beta)$$

Likelihood function $\ln L(\beta) = \sum \{y_i X_i' \beta - \exp(X_i' \beta) - \ln y_i\}$

Parameters β_p is estimated by k number of first order conditions as

$$\sum_{i=1}^n \{y_i X_i - \exp(X_i' \beta)\} X_i = 0$$

β_p is asymptotically normal with the sample covariance matrix

$$V[\hat{\beta}_p] = \left(\sum_{i=1}^n \lambda_i X_i X_i' \right)$$

Gauss-Newton or Newton-Raphson iterative algorithm is used to find unique parameter estimates.

IV. Tobit model

It is an extension of the probit model, named after James Tobin who adjusted Probit to handle truncation in a distribution. An observed sample may be truncated or censored, when observations below or above certain range have only one value; e.g. data is available only for those who buy a house but information does not exist for those who have not bought a house. We do not observe variables for people who have not bought a house.

$$Y_i = \begin{cases} \beta_0 + \beta_1 X_i + u_i & \text{if event occurs} \\ 0 & \text{otherwise} \end{cases}$$

Y_i is equal to $\beta_0 + \beta_1 X_i + u_i$ if the event is observed equal to zero if the event is not observed.

It is unscientific to estimate probability only with observed sample without worrying about the remaining observations in the truncated distribution. The Tobit model tries to correct this bias.

It is important to provide some background on the truncated and censored distributions in order to understand the Tobit model. Truncated and censored distribution is given by $f(x/x > a) = \frac{f(x)}{\Pr(x > a)}$. This normalization makes sure that area under the density function up to the censored point adds to 1.

$\int_a^{\infty} f(x/x > a) dx = 1$. The mean of truncated X is given by

$$E(x/x > a) = \int_a^{\infty} x f(x/x > a) dx = 1 \quad \text{and the mean of } X_{Trunc} \text{ is less than the mean of } X.$$

Normal distribution under truncation is modified as

$$f(x/x > a) = \frac{\frac{1}{\sigma} \phi\left(\frac{a - \mu}{\sigma}\right)}{1 - \Phi\left(\frac{a - \mu}{\sigma}\right)} = \frac{1}{\sigma} \frac{\phi(\alpha)}{1 - \Phi(\alpha)} \quad \text{where } \alpha = \frac{a - \mu}{\sigma}.$$

The mean of truncated variable is $E(x/x > a) = \mu + \sigma \lambda(\alpha)$ and its variance is $\text{var}(x/x > a) = \mu + \sigma^2(1 - \delta(\alpha))$, where $0 \leq \delta(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha] < 1$. Variance is also called a hazard function of this distribution. $\lambda(\alpha)$ is called inverse Mill's ratio. Here $0 < \delta(\alpha) < 1$ for all α .

$$\lambda(\alpha) = \begin{cases} \frac{\phi(\alpha)}{1 - \Phi(\alpha)} & \text{if } x > a \\ \frac{-\phi(\alpha)}{\Phi(\alpha)} & \text{if } x < a \end{cases}$$

Numerical Example

An affluent American makes 142,000 per year. Household survey had at least 100,000 income and only 2 percent belong to this category (from Greene (Chap 20). What is the mean and standard deviation of average income of all American households?

$$\text{Consider two equations: } E(x/x > a) = \mu + \sigma\lambda(\alpha); \quad \left. \begin{aligned} \lambda(\alpha) &= \frac{\phi(\alpha)}{(1 - \Phi(\alpha))} && \text{if } x > a \\ \lambda(\alpha) &= \frac{-\phi(\alpha)}{\Phi(\alpha)} && \text{if } x < a \end{aligned} \right\}$$

Assuming income are log normally distributed: $\text{LN}(100) = 4.602$; $\ln(142) = 4.956$;

$$\text{The standard normal: } \alpha = \frac{a - \mu}{\sigma}; \quad \Phi\left(\frac{a - \mu}{\sigma}\right) = 0.98;$$

$$\left(\frac{4.605 - \mu}{\sigma}\right) = \Phi^{-1}(0.98) = 2.054 \text{ [from the standard normal table } \Pr(Z \geq z)].$$

This yields: $\mu + 2.054\sigma = 4.605$ (1)

$$\lambda(\alpha) = \frac{\phi(\alpha)}{(1 - \Phi(\alpha))} = \frac{(2.054)}{1 - 0.98} = \frac{0.484}{0.02} = 2.420 \text{ [standard normal table } \Pr(Z \leq z)]$$

$$E(x/x > a) = \mu + \sigma\lambda(\alpha); \quad \mu + 2.420\sigma = 4.956 \text{ (2)}$$

Solving (1) and (2) yields $\sigma = 0.959$ and $\mu = 2.635$; and average income

$$\bar{Y} = e^{\frac{\mu + \sigma^2}{2}} = e^{2.635 + \frac{0.959^2}{2}} = 22.087$$

Truncated regression model

$$Y_i = \beta X_i + \varepsilon_i \quad \text{where } \varepsilon_i \sim N(0, \sigma^2) \quad (Y_i/X_i) \sim N(\beta_1 X_i, \sigma^2)$$

$$E(Y_i/X_i) = X_i\beta \text{ if truncated } (Y_i/y_i > a) = X_i\beta + \sigma\lambda_i(\alpha_i)$$

$$E(Y_i/y_i > a) = X_i\beta + E(\varepsilon_i/\varepsilon_i > a) > a - X_i\beta = X_i\beta + \sigma \frac{\phi\left(\frac{a - X_i\beta}{\sigma}\right)}{1 - \Phi\left(\frac{a - X_i\beta}{\sigma}\right)} = X_i\beta + \sigma\lambda_i(\alpha_i)$$

This is a non-linear function and the estimates of β and σ by the OLS technique is neither efficient nor consistent. Because the OLS ignores the

$$\lambda(\alpha) = \frac{\frac{1}{\sigma} \phi\left(\frac{a - X_i\beta}{\sigma}\right)}{1 - \Phi\left(\frac{a - X_i\beta}{\sigma}\right)} = \frac{1}{\sigma} \frac{1}{1 - \Phi(\alpha)} \phi(\alpha) \text{ part; } \lambda(\alpha) \text{ obviously depends on } X_i.$$

This is called inverse Mill's ratio. The unknown parameters β and σ can be estimated consistently by the Maximum likelihood technique as:

$$\log L = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum (Y_i - \beta X_i)^2 - \sum \log \left(1 - \Phi\left(\frac{a - \beta' X_i}{\sigma}\right)\right)$$

First order conditions for estimation of this ML are:

$$\frac{\partial \log L}{\partial \left(\frac{\beta}{\sigma^2} \right)} = \sum_{i=1}^n \left[\left(\frac{(Y_i - \beta X_i)}{\sigma^2} - \frac{\lambda_i}{\sigma} \right) X_i \right] = \sum_{i=1}^n g_i = 0$$

Maximise this likelihood by choosing β and σ iteratively taking the OLS estimates as the initial starting point.

Censored Distribution

Variable y_i^* is censored when it assumes only one value for all range of omitted part of the distribution. New variable is created by transforming the censored variable which results in a mixture of continuous and discrete parts as following

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0 \\ 0 & y_i^* \leq 0 \end{cases}$$

$$\text{If } y_i^* \sim N(\mu, \sigma^2); \Pr(y_i = 0) = \Pr(y_i^* \leq 0) = \Phi\left(-\frac{\mu}{\sigma}\right) = 1 - \Phi\left(\frac{\mu}{\sigma}\right)$$

$$\text{The distribution of the transformed variable is } g(y_i) = \begin{cases} N(\mu, \sigma^2) & \text{if } y_i^* > 0 \\ 1 - \Phi\left(\frac{\mu}{\sigma}\right) & y_i^* \leq 0 \end{cases}$$

Moments of the censored variable

$$E(y) = \Phi a + (1 - \Phi)(\mu + \sigma\lambda)$$

$$\text{var}(y) = \sigma^2(1 - \Phi) \left[(1 - \delta) + (\alpha - \lambda)^2 \Phi \right] \text{ where } \delta = \lambda^2 - \lambda\alpha; \lambda = \frac{\phi}{1 - \Phi};$$

$$E(y_i^* \leq 0) = \Phi(\alpha) = \Phi$$

Instead of scaling full probability is assigned to the censored region.

$$\text{When } a = 0 \text{ the mean simplifies to } E(y/a = 0) = \Phi\left(\frac{\mu}{\sigma}\right)(\mu + \sigma\lambda) \text{ with } \lambda = \frac{\phi(\mu/\sigma)}{\Phi(\mu/\sigma)}.$$

Tobit Model

$$Y_i = X_i\beta + \varepsilon_i \quad Y_i = \begin{cases} Y_i^* & \text{if } Y_i^* > 0 \\ 0 & Y_i^* \leq 0 \end{cases}$$

Calculation of mean function with censoring

$$E(Y_i/X_i) = \Phi\left(\frac{X_i\beta}{\sigma}\right)(X_i\beta + \sigma\lambda_i) \text{ with } \lambda_i = \frac{\phi(X_i\beta/\sigma)}{\Phi(X_i\beta/\sigma)}$$

$$E(Y_i/X_i, Y_i^* > 0) = X_i\beta + \sigma\lambda_i$$

$$E(Y_i^*/X_i) = X_i\beta$$

Marginal values:

$$\frac{\partial E(Y_i/X_i)}{\partial X_i} = \beta \Phi\left(\frac{X_i \beta}{\sigma}\right)$$

$$\frac{\partial E(Y_i/X_i, Y_i^* > 0)}{\partial X_i} = \beta \{ \Phi_i [1 - \lambda_i(\alpha_i) + \lambda_i] + \phi_i(\alpha + \lambda_i) \}; \quad \Phi_i = \Phi\left(\frac{\beta X_i}{\sigma}\right) = \Phi(\alpha_i)$$

$$\frac{\partial E(Y_i/X_i)}{\partial X_i} = \beta$$

The unknown parameters β and σ can be estimated consistently by the Maximum likelihood technique as:

$$\log L = \left[\sum_{Y_i^* > 0} -\frac{n}{2} \log 2\pi + \log \sigma^2 + \frac{(Y_i - \beta X_i)^2}{\sigma^2} \right]_{\text{Uncensored}} + \sum_{Y_i^* = 0} \left(1 - \Phi\left(\frac{X_i \beta}{\sigma}\right) \right)_{\text{Censored}}$$

Uncensored part

Censored part

Use the OLS estimates as the starting values.

The two step estimation procedure proceeds as following:

First construct the index variable $I_i = \int Y_i^* \text{ if } Y_i^* > 0$
 0 Otherwise

$\Pr(I_i = 1) = 1 - \Phi_i = p_i$ and $\Pr(I_i = 0) = \Phi_i = 1 - p_i$

Then the log likelihood

$$L = \prod_{I_i=1} P_i \prod_{I_i=0} (1 - P_i)$$

Using Probit and MLE find estimates of parameters β and σ and λ

$$\hat{\lambda}_i = \frac{\phi(X_i \hat{\beta} / \hat{\sigma})}{\Phi(X_i \hat{\beta} / \hat{\sigma})}$$

Secondly apply OLS at the second stage:

$Y_i = X_i \beta + \sigma \lambda_i + w_i$; where w_i is the composite error.

$$Y_i = Z_i \gamma + w_i$$

where $Z_i = (X_i, \lambda_i)$ and $\gamma = \begin{pmatrix} \beta \\ \sigma \end{pmatrix}$

$$\hat{\gamma} = (Z'Z)^{-1} Z'X$$

$$\text{var}(\hat{\gamma}_{2\text{stage}}) = (Z'Z)^{-1} Z'AZ(Z'Z)^{-1} \text{ and } A_i = Y_i - X_i \beta - \sigma \lambda_i$$

[$E(Y_i) = \hat{\Phi} X_i \beta + \sigma \hat{\phi}_i$ or $E(Y_i) = \hat{\Phi} X_i \beta + \sigma \phi_i$ which one of these two better not clear (Amimiya (1984; JEL)].

Variations of censoring

$$Y_i = \begin{cases} X_i \beta + \varepsilon_i & \text{if } Y_i^* > c_i \\ c_i & \text{Otherwise} \end{cases}$$

c_i can be stochastic of other variables.

$$\text{define } Y_i = (y_i, c_i); X_i^* = \begin{bmatrix} X_i \\ c_i \end{bmatrix}; \beta^* = \begin{bmatrix} \beta \\ -1 \end{bmatrix}$$

Similarly there are other upper limit distributions

$$Y_i = \begin{cases} X_i \beta + \varepsilon_i & \text{if } RHS < 0 \\ 0 & \text{Otherwise} \end{cases}$$

Specification issues:

Heteroscedasticity, Autocorrelation, lag dependent variable, normality of errors, use of other distributions such as Weibul, Exponential, log normal, least absolute deviation, Box Cox transformation or other robust methods.

ML estimates are inconsistent with heteroscedasticity, consistent but not efficient with autocorrelation. Methods of moments are more robust than ML method .

Misspecification test

$$\sigma_i^2 = (\gamma + \delta Z_i)^2$$

$$\log L = \sum_{Y_i^* > 0} -\frac{n}{2} \log 2\pi + \log \sigma^2 + \frac{(Y_i - \beta X_i)^2}{\sigma^2} + \sum_{Y_i^* > 0} \left(1 - \Phi \left(\frac{X_i \beta}{\sigma} \right) \right) \quad H_0 : \delta = 0$$

$$\begin{bmatrix} E_{MM} & E_{ML} \\ E_{ML} & E_{ML} \end{bmatrix} \begin{bmatrix} V_{MM} & V_{ML} \\ V_{ML} & V_{ML} \end{bmatrix} \begin{bmatrix} E_{MM} & E_{ML} \\ E_{ML} & E_{ML} \end{bmatrix} \sim \chi_{df=k}^2$$

If significant $\chi_{df=k}^2$ reject H_0 that model is mis-specified.

Conditional moment tests (Pagan and Vella)

Hetero; omitted variable

1. $E[Z_i(Y_i - X_i \beta)] = 0$
2. $E[Z_i(Y_i - X_i \beta)^2 - \sigma^2] = 0$

Population

3. $E[Z_i(Y_i - X_i \beta)]^3 = 0$ -Skewness test and $E[Z_i(Y_i - X_i \beta)]^4 - 3\sigma^4 = 0$ Kurtosis test;

Sample counterparts of the moment restrictions in (1) to (3) are

- 1) for hetero $\gamma_1 = \frac{1}{n} \sum Z_i (e_i^2 - s_i^2) e_i = Y_i - X_i \beta_{OLS}$

- 2) Omitted variables $\gamma_2 = \frac{1}{n} \sum Z_i e_i^2$

- 3) $\gamma_3 = \frac{1}{n} \left[\frac{\sum e_i^3}{\sum e_i^4 - 3s^4} \right]$

Sample counterparts

1. $m_i = [Z_i(Y_i - X_i \beta)]$

$$2. m_i = [Z_i(Y_i - X_i\beta)^2 - \sigma^2]$$

Population

$$3. m_i = [Z_i(Y_i - X_i\beta)]^3 = 0 ; m_i = [Z_i(Y_i - X_i\beta)]^4 - 3\sigma^4 = 0 \text{ Kurtosis test;}$$

$$\gamma_j = \frac{1}{n} \sum_i m_i(Y_i X_i, Z_i, \hat{\beta}, \hat{\sigma}) = \frac{1}{n} \sum_i m_i$$

$$H_0 : E(\gamma_i) = 0 ; E(m_i) = 0 ; \sqrt{n}\gamma_i = N(0, \Sigma)$$

misspecification test: $n\gamma_j \hat{\Sigma}^{-1} \gamma_j \sim \chi_{df=k}^2$; this needs sample estimates.

Two limit Tobit

$$Y_i = X_i\beta + \varepsilon_i \quad Y_i = \begin{cases} Y_i^* & \text{if } L_1 < Y_i^* < L_2 \\ L_{1,i} & \text{if } Y_i^* < L_1 \\ L_{2,i} & \text{if } Y_i^* \geq L_2 \end{cases}$$

Formulate a maximum likelihood function

$$L = \prod_{Y_i=L_1} \Phi\left(\frac{L_{1,i} - X_i\beta}{\sigma}\right) \prod_{Y_i=Y_i^*} \Phi\left(\frac{Y_i - X_i\beta}{\sigma}\right) \prod_{Y_i=L_{2,i}} \Phi\left(\frac{L_{2,i} - X_i\beta}{\sigma}\right)$$

For instance think of a minimum, partial or maximum coverage of insurance.

Expected values

$$E(Y_i/X_i, L_1 < Y_i^* < L_2) = X_i\beta + \frac{\sigma(\Phi_{1,i} - \Phi_{2,i})}{\Phi_{2,i} - \Phi_{1,i}}$$

$$E(Y_i/X_i) = \Phi_{1,i}L_{1,i} + X_i\beta(\Phi_{2,i} - \Phi_{1,i}) + \sigma(\phi_{1,i} - \phi_{2,i}) + (1 - \Phi_{2,i})L_{2,i}$$

$$\Phi_{1,i} = \Phi\left(\frac{L_{1,i} - X_i\beta}{\sigma}\right); \Phi_{2,i} = \Phi\left(\frac{L_{2,i} - X_i\beta}{\sigma}\right)$$

$$L = \prod_{n1} \Phi_{1,i} \prod_{n2} (\Phi_{2,i} - \Phi_{1,i}) \prod_{n3} (1 - \Phi_{2,i})$$

$$\log L = N_2 \log \rho + N \log \sigma + \sum_{i=1}^N (Y_i - \beta X_i)^2 - \sum \log[\rho(1 - \rho)\Phi_i]$$

If ρ known estimate β_{ML} and σ_{ML} if not β_{ML} and σ_{ML} and test for $\rho = 1$

$$E(Y_i/X_i; I_i = 1) = X_i\beta + E(\varepsilon_i/I_i = 1) = X_i\beta + \sigma \frac{\phi_i}{\Phi_i}$$

$$E(Y_i/X_i; I_i = 0) = X_i\beta + \sigma \frac{\phi_i}{1 - \Phi_i}$$

$$E(Y_i/X_i) = X_i\beta = \frac{\sigma\rho_1\Phi_i - \rho_2\Phi_i}{\rho_1\Phi_i - \rho_2\Phi_i}$$

Heckman's Model

Inference population when the sample is non-random and some observations are omitted causes a sample selection bias. Heckman's procedure is to microeconometrics as is the unit root for time series data. When the sample selection is not corrected inference drawn from the regression analysis is not efficient or robust.

$$Y_{1,i} = X_{1,i}\beta + \varepsilon_{1,i} \quad \text{if } Y_{1,i} > Y_{2,i}$$

$$Y_{2,i} = X_{2,i}\beta + \varepsilon_{2,i}$$

Both $Y_{1,i}$ and $Y_{2,i}$ are stochastic. For instance, if $Y_{1,i}$ is market wage and $Y_{2,i}$ is the reservation wage; an individual works only when $Y_{1,i} > Y_{2,i}$.

$$\begin{pmatrix} \varepsilon_{1,i} \\ \varepsilon_{2,i} \end{pmatrix} \sim IN(\underline{0}, \Sigma) \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_2^2 \end{bmatrix}$$

At least one more variable in $X_{2,i}$ than in $X_{1,i}$; if $\sigma_{1,2} = 0$ it is a regular Tobit.

But the sample selection problem arise when $\sigma_{1,2} \neq 0$.

$$Y_{1,i} > Y_{2,i} \quad \text{implies} \quad \varepsilon_{1,i} = Y_{1,i} - X_{1,i}\beta$$

$$\varepsilon_{2,i} < Y_{2,i} - X_{2,i}\beta$$

$$\text{Joint density of } (\varepsilon_{1,i}, \varepsilon_{2,i}) = f(\varepsilon_{1,i}, \varepsilon_{2,i}) = g(\varepsilon_{1,i})h(\varepsilon_{2,i}/\varepsilon_{1,i})$$

For some individuals you do not observe $Y_{1,i}$ because $Y_{1,i} < Y_{2,i}$.

$$Y_{1,i} < Y_{2,i} = \varepsilon_{1,i} - \varepsilon_{2,i} < X_{2,i}\beta - X_{1,i}\beta$$

$$(\varepsilon_{1,i}, \varepsilon_{2,i}) = IN(\underline{0}, \sigma^2); \quad \sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2}$$

$$\log L = -N_1 \log \sigma^2 - \frac{1}{2\sigma_1^2} \sum_{i=1}^N (Y_{1,i} - X_{1,i}\beta)^2 + \sum_{\text{Instrumnt}} \log \Phi_i(w) + \sum_0 \log \left[\frac{(X_{2,i}\beta_2 - X_{1,i}\beta_1)}{\sigma} \right]_{\text{Unobsorved}}$$

$$w = \frac{Y_{2,i} - X_{2,i}\beta_2}{\sigma_{2,1}} - \frac{\sigma_{1,2}}{\sigma_1^2} (Y_{1,i} - X_{1,i}\beta_1)$$

$$W = \beta_0 + X_{1,i}\beta + \varepsilon_{1,i}$$

$$W_R = \gamma_0 + \gamma_1 H + \gamma_2 Z + \varepsilon_{2,i}$$

X and Z are exogenous variables and H and W are endogenous. If $H = 0$ $W_R > W$.

$$H = \frac{\beta_0 + X_{1,i}\beta - \gamma_0 - \gamma_2 Z}{\gamma_1} + \frac{\varepsilon_{2,i} - \varepsilon_{1,i}}{\gamma_1} = X\delta + V$$

$$\left. \begin{aligned} W &= \beta_0 + X_{1,i}\beta + \varepsilon_{1,i} \\ H &= X\delta + V \end{aligned} \right\} \quad \text{if } H > 0$$

$$W = H = 0 \quad \text{Otherwise}$$

$$\Pr(H = 0) = \Pr(V \leq -R\delta) = pr \left(\frac{\varepsilon_{2,i} - \varepsilon_{1,i}}{\gamma_1} \leq \frac{\gamma_0 + \gamma_2 Z - \beta_0 - X_{1,i}\beta}{\gamma_1} \right)$$

$$\Pr(\varepsilon_{2,i} - \varepsilon_{1,i} \leq (\gamma_0 + \gamma_2 Z - \beta_0 - X_{1,i} \beta)) = \Phi(\Delta)$$

$$\Delta = \frac{\gamma_0 + \gamma_2 Z - \beta_0 - X_{1,i} \beta}{\sigma} ; \sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2}$$

$$L = \prod_{H>0} WH \prod_{H=0} (\Delta)$$

$$E(W/H > 0) = \beta_0 + X_{1,i} \beta + (\varepsilon_{1,i}/H > 0) = \beta_0 + X_{1,i} \beta + \frac{\alpha \phi(\Delta)}{1 - \Phi(\Delta)} = \beta_0 + X_{1,i} \beta + \alpha \lambda$$

$$E(W/H > 0) = \beta_0 + X_{1,i} \beta + \alpha \lambda = \beta_0 + X_{1,i} \beta + \sigma_{1,2} \frac{\phi\left(\frac{\gamma_0 + \gamma_2 Z - \beta_0 - X_{1,i} \beta}{\sigma}\right)}{1 - \Phi\left(\frac{\gamma_0 + \gamma_2 Z - \beta_0 - X_{1,i} \beta}{\sigma}\right)}$$

Thus the sample selection bias is due to the **Heckman's Lambda** term.

$$\begin{aligned} E(\varepsilon_{2,i}/H > 0) &= E(\varepsilon_{2,i}/V > -RD) = E(\varepsilon_{2,i}/\varepsilon_{2,i} - \varepsilon_{1,i} > \gamma_0 + \gamma_2 Z - \beta_0 - X_{1,i} \beta) = \\ &= \frac{\sigma_2^2 - 2\sigma_{1,2}}{\sigma} \frac{\phi(\Delta)}{1 - \Phi(\Delta)} = \frac{\alpha \phi(\Delta)}{1 - \Phi(\Delta)} \end{aligned}$$

To estimate $E(W/H > 0) = \beta_0 + X_{1,i} \beta + \alpha \lambda$

Use probit to estimate $\hat{\lambda} \quad I_i = \int_0^1 \text{if working}$
 $0 \quad \text{Otherwise}$

and replace it in $W_i = \beta_0 + X_{1,i} \beta + \alpha \hat{\lambda} + \eta_i$

$$\Pr(I_i = 1) = \Pr(H > 0) = 1 - \Phi(\Delta)$$

$$\Pr(V > -RD) = \Pr(\varepsilon_{2,i} - \varepsilon_{1,i} > \gamma_0 + \gamma_2 Z - \beta_0 - X_{1,i} \beta) \quad \Pr(I_i = 0) = \phi(\Delta)$$

Heckman's Tobit then:

$$L = \prod_{H>0} (1 - \Phi(\Delta)) \prod_{H \leq 0} \Phi(\Delta) \text{ by MLE get } \Delta \text{ and } \hat{\lambda} = \frac{\phi(\hat{\Delta})}{1 - \Phi(\hat{\Delta})}$$

$$H = \frac{\beta_0 + X_{1,i} \beta - \gamma_0 - \gamma_2 Z}{\gamma_1} + \frac{\varepsilon_{2,i} - \varepsilon_{1,i}}{\gamma_1} = \hat{\Delta}$$

Apply OLS to $W_i = \beta_0 + X_{1,i} \beta + \alpha \hat{\lambda} + \eta_i$

Simultaneous equation method

$$\left. \begin{aligned} Y_1^* &= \gamma_1 Y_2^* + \beta_1 X_1 + \varepsilon_1 \\ Y_2^* &= \gamma_2 Y_1^* + \beta_2 X_2 + \varepsilon_2 \end{aligned} \right\}$$

Consider two cases; in case 1 Y_1 is observed Y_2 is censored. In case 2 both Y_1 and Y_2 are censored.

Case 1

$$Y_1 = \Pi_1 X + V_1$$

$$Y_2 = \Pi_2 X + V_2$$

where $X = (X_1, X_2)$; First apply Tobit to the second equation and substitute \hat{Y}_2 for Y_2^* in equation 1 and then apply OLS to estimate Π_1 and then \hat{Y}_1 for Y_1^* in equation 2. Then retrieve the structural coefficients.

Switching regressions: Regime selection and disequilibrium

$$\text{Regime 1: } Y_{1,i} = X_{1,i}\beta + u_{1,i} \quad \text{observe } Y_i = \begin{cases} Y_{1,i} & \text{if } I_i = 1 \\ Y_{2,i} & \text{if } I_i = 0 \end{cases}$$

$$\text{Regime 2: } Y_{2,i} = X_{2,i}\beta + u_{2,i}$$

$$\text{Participation mechanism } I^* = Z\alpha - \varepsilon$$

$$\text{and } (I^*, I) \quad I_i = \begin{cases} 1 & \text{if } I_i^* = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{pmatrix} u_{1,i} \\ u_{2,i} \\ \varepsilon \end{pmatrix} \sim IN(0, \Sigma) \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,\varepsilon} \\ & \sigma_2^2 & \sigma_{2,\varepsilon} \\ & & \sigma_\varepsilon^2 \end{bmatrix}$$

$\sigma_{1,\varepsilon} = 0$ and $\sigma_{2,\varepsilon} = 0$ implies exogenous switching otherwise it will be endogenous switching. Consider switching where sample separation is known:

1. Y observed in regime 1 and regime 2; but the dependent variable is observed only in regime 1 or regime 2; when sample separation is unknown you do not know whether Y belongs to regime 1 or regime 2.

Example

If $W_o > W_R$, this individual works; $W_o < W_R$ does not work.

Distribution of wage observed is truncated by self selection.

$$W_o = X\beta_1 + u_1$$

$$W_R = X\beta_2 + u_2$$

$$W = \begin{cases} W_o & \text{if } W_o > W_R \\ 0 & \text{otherwise} \end{cases}$$

Selectivity model $Y_i = X_i\beta + \varepsilon_i$ is model of interest

$$\text{Selection criteria } Z_i = \begin{cases} 1 & \text{if } Z_i^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Error structure } (\varepsilon, u) = (0, 0, \sigma_\varepsilon^2, \sigma_u^2, 2\sigma_\varepsilon\sigma_u, \rho)$$

$$E(Y_i/Y_i \text{ observed}) = E(Y_i/Z_i = 1) = E(Y_i/Z_i^* > 1) = E(Y_i/u_i > -W_R)$$

$$= X_i\beta + \rho\sigma_\varepsilon\lambda_i(\alpha_i) = X_i\beta + \theta\lambda_i$$

$$\alpha_i = -\frac{W_{i,R}}{\sigma_u} \lambda_i(\alpha_i) = \phi\left(\frac{W_{i,R}/\sigma_u}{W_{i,R}/\sigma_\varepsilon}\right)$$

1. OLS is biased and inconsistent for estimation of β
2. SSB test for $\theta = 0$
3. OLS consistent with λ_i but inefficient because of heterogeneity
4. Marginal effect varies with X_i

$$\frac{\partial E(Y_i/Y_i \text{ observed})}{\partial X_{i,k}} = \beta_k - \gamma_k \left(\rho \frac{\sigma_\varepsilon}{\sigma_u} \right) \delta_i \text{ where } \delta_i = \lambda_i^2 + \alpha \lambda_i$$

ML is optimal but two step is convenient.

Two step estimation for the selectivity model; first Probit and then ML.

$$L = \Pi \Pr(Z_i = 1) \Pi(Z_i = 0)$$

$$\text{where } \Pr(Z_i = 1) = \Pr\left(\frac{u}{\sigma_u} > \frac{W_{i,R}}{\sigma_u}\right) = \Phi\left(\frac{W_{i,R}}{\sigma_u}\right) = \Phi_i$$

$$L = \Pi \Phi_i \Pi(1 - \Phi_i) \text{ by MLE } \lambda_i = \frac{\phi(W_{i,R}/\sigma_u)}{\Phi(W_{i,R}/\sigma_u)}; \hat{\delta}_i = \hat{\lambda}(\lambda - \hat{\alpha}) = \frac{W\hat{\alpha}}{\sigma_u}$$

with observed Y regress Y on X and $\hat{\lambda}$ for $\hat{\beta}$, $\hat{\theta}$ and $\hat{\rho} = \frac{\hat{\theta}^2}{\sigma_\varepsilon^2}$

$$\sigma_\varepsilon^2 = \frac{\varepsilon' \varepsilon}{n} + \delta \hat{\theta}^2 \quad \delta = \frac{1}{n} \sum \hat{\delta}_i$$

$$\text{var}\begin{pmatrix} \hat{\beta} \\ \hat{\theta} \end{pmatrix} = \hat{\sigma}_\varepsilon^2 (X' X^*)^{-1} (X^{-1} (1 - \rho^2 \Delta) X^* + \theta) (X' X^*)$$

$X^* = (X, \hat{\lambda})$; $\theta = \rho^2 F \hat{V} F$ - asymptotic variance; $F = X^* \Delta W$ - set of exogenous variables.

Two step procedure when sample is known

Regime 1: $\theta_t = X_{1,i} \beta + u_{1,i}$ iff $X_{1,i} \beta + u_{1,i} < X_{2,i} \beta + u_{2,i}$

Regime 2: $\theta_t = X_{2,i} \beta + u_{2,i}$ iff $X_{1,i} \beta + u_{1,i} > X_{2,i} \beta + u_{2,i}$

Criterion $Z\gamma \geq u_t$ in regimes and if $Z\gamma < u_t$ in regime 2.

$$Z = \frac{X_{1,i} \beta - X_{2,i} \beta \gamma_0 - \gamma_2 Z}{\sigma}; u = \frac{u_{2,i} - u_{1,i}}{\sigma}; \sigma_1^2 - \sigma_2^2 = \text{var}(u_{2,i} - u_{1,i})$$

$$\begin{pmatrix} u_{1,i} \\ u_{2,i} \\ \varepsilon \end{pmatrix} \sim IN(\underline{0}, \Sigma) \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,\varepsilon} \\ & \sigma_2^2 & \sigma_{2,\varepsilon} \\ & & \sigma_\varepsilon^2 \end{bmatrix}$$

Regime 1: $\theta_t = X_{1,i}\beta - \sigma \frac{\phi(Z\gamma)}{\Phi(Z\gamma)} + u_{1,i}$

Regime 2: $\theta_t = X_{2,i}\beta + \sigma \frac{\phi(Z\gamma)}{1 - \Phi(Z\gamma)} + u_{2,i}$

Calculate $\frac{\phi}{\Phi}$ using Probit; where $L = \prod_{R2} \Pr(Z\gamma \geq 1) \prod_{R1} \Pr(Z\gamma < 0)$

Then apply OLS to estimate $\hat{\beta}_1 \hat{\beta}_2 \sigma_1^2 \sigma_{1,2} \sigma_{1,\epsilon}$

Disequilibrium models

Consider partial equilibrium model

$$y_t = y_{t-1} + \lambda(y_t^* - y_{t-1})$$

y_t^* is desired value; $(y_t^* - y_{t-1})$ measures the disequilibrium $\lambda < 1$ is the adjustment towards equilibrium. Loss function including cost of disequilibrium and the cost of adjustment.

$$L = \alpha(y_t^* - y_t)^2 + \beta(y_t - y_{t-1})^2$$

with minimisation of L $y_t = y_{t-1} + \frac{\alpha}{\alpha + \beta}(y_t^* - y_{t-1})$ and $\lambda = \frac{\alpha}{\alpha + \beta}$ the adjustment coefficient if function of disequilibrium as well as adjustment costs.

Take a demand and supply model

Demand : $D_t = \alpha_1 P_t + X_{1,t} \beta_1 + u_{1,t}$

Supply: $S_t = \alpha_2 P_t + X_{2,t} \beta_2 + u_{2,t}$

Equilibrium: $Q_t = D_t = S_t$

Disequilibrium : $Q_t = \begin{cases} D_t > S_t & Q_t \text{ on } S \\ D_t < S_t & Q_t \text{ on } D \end{cases}$

Let λ_t be probability of observing D; $\Pr(D_t < S_t) = \Pr(u_{1,t} - u_{2,t} < X_{2,t} \beta_2 - X_{1,t} \beta_1)$ and $(1 - \lambda_t)$ be the probability of observing $\Pr(S_t < D_t)$. Then the joint density

$$g(D, S)$$

$$h(Q_t) = \lambda_t h(Q_t / Q_t = D_t) + (1 - \lambda_t) h(Q_t / Q_t = S_t) = \int_{Q_t}^{\infty} g(Q_t, S_t) dS_t + \int_{Q_t}^{\infty} g(Q_t, D_t) dD_t$$

Log likelihood: $L = \Pi h(Q_t)$ but this may be unbounded and restricted.

When sample separation in known $g_{1,t} = \int_{Q_t}^{\infty} g(Q_t, S_t) dS_t / (1 - \lambda_t)$

$$g_{2,t} = \int_{Q_t}^{\infty} g(Q_t, D_t) dD_t / \lambda_t$$

Likelihood function:

$$L = \prod_{\Delta P < 0} g_{1,t} \prod_{\Delta P > 0} g_{2,t} \prod_{\Delta P < 0} \lambda_t \prod_{\Delta P > 0} (1 - \lambda_t)$$

Price directions

$$\Delta P > 0 \begin{cases} D_t > S_t & Q_t \text{ on } S \\ D_t < S_t & Q_t \text{ on } D \end{cases}$$

References:

Discrete Choice Laboratory.

Arulampalam W., Booth A.L., (1998), "Learning and Earning: Do Multiple Training Events Pay?", *Journal of Economic Literature*.

Bhattacharai (1995) A Notes on Cross Section Analysis, based on Lecture notes, mimeo.

Blundell R W and R J. Smith (1989) Estimation in a class of simultaneous equation limited dependent variable models, *Review of Economic Studies*, 56:37-38.

Chesher A (1984) Improving the efficiency of Probit estimators, *Review of Economic Studies*, 66:3:523-527.

Greene W. (2000) *Econometric Analysis*, Prentice Hall.

Greene W.H. (1998) LIMDEP Version 7: User Manual, Econometric Software Inc.

Hausman J.A., (1978), "Specification Tests in Econometrics", *Econometrica*, Vol. 46, No. 6, pp.1251-1271.

Heckman J. J., (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, Vol. 47, No. 1, pp.153-161.

Imbens G. W. and T Lancaster (1994) Combining Micro and Macro Data in Microeconomic Models, *Review of Economic Studies*, 61:4:655-680.

Judge G.G., W.E. Griffiths, R. C. Hill, H. Lutkepohl and T.C. Lee(1990) *Theory and Practice of Econometrics*, Wiley.

Keifer N (1988) Economic duration data and hazard functions, *Journal of Economic Literature*, 26:647-679.

Lancaster T and A Chesher (1983) The Estimation of Models of Labour Market Behaviour *Review of Economic Studies*, 50:4:609-624.

Maddala G.S. (1983) *Limited-dependent and qualitative variables in econometrics*, Cambridge University Press, New York.

McFadden D.L. (1974) Conditional logit analysis of qualitative choice behaviour in P.Zarembka (ed.) *Frontiers in Econometrics*, New York, Academic Press.

Staigler D., Stock J. H., (1997), "Instrumental Variables Regression with Weak Instruments", *Econometrica*, Vol. 65, No. 3, pp.557-586.

Theil H. (1969) A multinomial extension of the linear logit model, *International Economic Review*, 10:251-9.

Tobin J (1958) Estimation of relationship for limited dependent variables, *Econometrica*, 26:24-36.

Verbeek M. (2004) *A Guide to Modern Econometrics*, Wiley.

Wooldridge J. M. (2002) *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

```

/*=====
Greene Example 19.1. Probability Models
*/=====
Read ; Nobs = 32 ; nvar = 5 ;          9  3.03  25  0  0          21  2.06  22  1  0
Names = 1 $                            10  3.92  29  0  1          22  3.62  28  1  1
OBS   GPA   TUCE  PSI                    11  2.63  20  0  0          23  2.89  14  1  0
GRADE                                     12  3.32  23  0  0          24  3.51  26  1  0
1   2.66  20  0  0                        13  3.57  23  0  0          25  3.54  24  1  1
2   2.89  22  0  0                        14  3.26  25  0  1          26  2.83  27  1  1
3   3.28  24  0  0                        15  3.53  26  0  0          27  3.39  17  1  1
4   2.92  12  0  0                        16  2.74  19  0  0          28  2.67  24  1  0
5   4.00  21  0  1                        17  2.75  25  0  0          29  3.65  21  1  1
6   2.86  17  0  0                        18  2.83  19  0  0          30  4.00  23  1  1
7   2.76  17  0  0                        19  3.12  23  1  0          31  3.10  21  1  0
8   2.87  21  0  0                        20  3.16  25  1  1          32  2.39  19  1  1
?
Namelist ; X = One,GPA,TUCE,PSI $
Matrix ; xbar = mean(x) $
Regress ; Lhs = Grade ; Rhs = X $
Calc ; List ; Scale = 1.0 $
Matrix ; List ; ME = Scale * b $
Probit ; Lhs = Grade ; Rhs = X ; Marginal Effects $
Calc ; List ; Scale = N01(b'xbar) $
Matrix ; List ; ME = Scale * b $
Logit ; Lhs = Grade ; Rhs = X ; marginal Effects $
Calc ; List ; Scale = LGD(b'xbar) $
Matrix ; List ; ME = Scale * b $
Create ; d0 = 1-Grade ; d1 = Grade $
Maximize ; Fcn = -d0*exp(b1'x) + d1*log(1-exp(-exp(b1'x)))
; Labels = b1,b2,b3,b4
; Start = b $
Calc ; List ; Scale = exp(-exp(b'xbar))*exp(b'xbar) $
Matrix ; List ; ME = Scale * b $
/*
+-----+
| Ordinary least squares regression Weighting variable = none |
| Dep. var. = GRADE Mean= .3437500000 , S.D.= .4825587044 |
| Model size: Observations = 32, Parameters = 4, Deg.Fr.= 28 |
+-----+
+-----+-----+-----+-----+-----+-----+
|Variable| Coefficient | Standard Error |t-ratio|P[|T|>t]| Mean of X|
+-----+-----+-----+-----+-----+-----+

```

Constant -1.498017120 .52388862 -2.859 .0079
 GPA .4638516793 .16195635 2.864 .0078 3.1171875
 TUCE .1049512224E-01 .19482854E-01 .539 .5944 21.937500
 PSI .3785547879 .13917274 2.720 .0111 .43750000
 SCALE = .1000000000000000D+01

Matrix ME has 4 rows and 1 columns.

```

+-----+
1|-.1498017D+01
2| .4638517D+00
3| .1049512D-01
4| .3785548D+00

```

```

+-----+
| Binomial Probit Model          |
| Number of observations         32 |
| Iterations completed           6 |
| Log likelihood function        -12.81880 |
| Restricted log likelihood       -20.59173 |
| Chi-squared                    15.54585 |
| Degrees of freedom              3 |
| Significance level              .1404896E-02 |
+-----+

```

```

+-----+-----+-----+-----+-----+
| Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
+-----+-----+-----+-----+-----+

```

Index function for probability

Constant -7.452319647 2.5424723 -2.931 .0034
 GPA 1.625810039 .69388249 2.343 .0191 3.1171875
 TUCE .5172894549E-01 .83890261E-01 .617 .5375 21.937500
 PSI 1.426332342 .59503790 2.397 .0165 .43750000

```

+-----+
| Partial derivatives of E[y] = F[*] with |
| respect to the vector of characteristics. |
| They are computed at the means of the Xs. |
+-----+

```

```

+-----+-----+-----+-----+-----+
| Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
+-----+-----+-----+-----+-----+

```

Index function for probability

Constant -2.444733653 .75885194 -3.222 .0013
 GPA .5333470255 .23246407 2.294 .0218 3.1171875
 TUCE .1696968191E-01 .27119788E-01 .626 .5315 21.937500
 PSI .4679083617 .18764238 2.494 .0126 .43750000
 SCALE = .32805002591068580D+00

Matrix ME has 4 rows and 1 columns.

```

+-----+
1|-.2444734D+01

```

```

2| .5333470D+00
3| .1696968D-01
4| .4679084D+00
+-----+
| Multinomial Logit Model          |
| Dependent variable      GRADE    |
| Number of observations      32    |
| Iterations completed      6      |
| Log likelihood function  -12.88963 |
| Restricted log likelihood  -20.59173 |
| Chi-squared              15.40419 |
| Degrees of freedom        3      |
| Significance level        .1501878E-02 |
+-----+
+-----+-----+-----+-----+-----+
| Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
+-----+-----+-----+-----+-----+
          Characteristics in numerator of Prob[Y = 1]
Constant -13.02134648  4.9313241  -2.641  .0083
GPA      2.826112525   1.2629411  2.238  .0252  3.1171875
TUCE     .9515765670E-01 .14155420   .672  .5014  21.937500
PSI      2.378687596   1.0645642  2.234  .0255  .43750000
+-----+
| Partial derivatives of probabilities with |
| respect to the vector of characteristics. |
| They are computed at the means of the Xs. |
| Observations used for means are All Obs. |
+-----+
+-----+-----+-----+-----+-----+
| Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
+-----+-----+-----+-----+-----+
          Marginal effects on Prob[Y = 1]
Constant -2.459760743  .81771031  -3.008  .0026
GPA      .5338588183    .23703797   2.252  .0243  3.1171875
TUCE     .1797548884E-01 .26236909E-01 .685  .4933  21.937500
PSI      .4493392735     .19676264   2.284  .0224  .43750000
SCALE = .18890218048721940D+00
Matrix ME has 4 rows and 1 columns.
+-----+
1| -.2459761D+01
2| .5338588D+00
3| .1797549D-01
4| .4493393D+00
+-----+
| User Defined Optimization          |
| Number of observations      32    |

```

```
| Iterations completed      10 |
| Log likelihood function  -13.00800 |
```

```
+-----+
+-----+-----+-----+-----+
| Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
+-----+-----+-----+-----+
B1  -10.03142629  3.4448183  -2.912  .0036
B2   2.293553481  .99127580  2.314  .0207
B3   .4115615961E-01  .12327466  .334  .7385
B4   1.562276316  .85886363  1.819  .0689

SCALE = .20817920176287030D+00
Matrix ME has 4 rows and 1 columns.
```

```
+-----+
1| -.2088334D+01
2| .4774701D+00
3| .8567856D-02
4| .3252334D+00
```

```
*/
--> Read ; Nobs = 32 ; nvar = 5 ; Names = 1 $
--> Namelist ; X = One,GPA,TUCE,PSI $
--> Matrix ; xbar = mean(x) $
--> Regress ; Lhs = Grade ; Rhs = X $
```

```
+-----+
| Ordinary least squares regression Weighting variable = none
| Dep. var. = GRADE Mean= .3437500000 , S.D.= .4825587044
| Model size: Observations = 32, Parameters = 4, Deg.Fr.= 28
| Residuals: Sum of squares= 4.216473951 , Std.Dev.= .38806
| Fit: R-squared= .415900, Adjusted R-squared = .35332
| Model test: F[ 3, 28] = 6.65, Prob value = .00157
| Diagnostic: Log-L = -12.9782, Restricted(b=0) Log-L = -21.5812
| LogAmemiyaPrCrt.= -1.775, Akaike Info. Crt.= 1.061
| Autocorrel: Durbin-Watson Statistic = 2.34645, Rho = -.17322
+-----+
```

```
+-----+-----+-----+-----+-----+
| Variable | Coefficient | Standard Error | t-ratio | P[|T|>t] | Mean of X |
+-----+-----+-----+-----+-----+
Constant -1.498017120 .52388862 -2.859 .0079
GPA .4638516793 .16195635 2.864 .0078 3.1171875
TUCE .1049512224E-01 .19482854E-01 .539 .5944 21.937500
PSI .3785547879 .13917274 2.720 .0111 .43750000
```

Matrix: Las
[4,4]

```
--> Calc ; List ; Scale = 1.0 $
SCALE = .10000000000000000D+01
--> Matrix ; List ; ME = Scale * b $
```

Matrix ME has 4 rows and 1 columns.

```
1
+-----+
1| -.1498017D+01
2| .4638517D+00
3| .1049512D-01
4| .3785548D+00
```

```
--> Probit ; Lhs = Grade ; Rhs = X ; Marginal Effects $
```

```
+-----+
| Dependent variable is binary, y=0 or y not equal 0
| Ordinary least squares regression Weighting variable = none
| Dep. var. = GRADE Mean= .3437500000 , S.D.= .4825587044
| Model size: Observations = 32, Parameters = 4, Deg.Fr.= 28
| Residuals: Sum of squares= 4.216473951 , Std.Dev.= .38806
| Fit: R-squared= .415900, Adjusted R-squared = .35332
| Model test: F[ 3, 28] = 6.65, Prob value = .00157
| Diagnostic: Log-L = -12.9782, Restricted(b=0) Log-L = -21.5812
| LogAmemiyaPrCrt.= -1.775, Akaike Info. Crt.= 1.061
+-----+
```

```

+-----+
+-----+-----+-----+-----+-----+-----+
|Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
+-----+-----+-----+-----+-----+-----+
Constant -1.498017120      .52388862      -2.859      .0042
GPA      .4638516793      .16195635      2.864      .0042      3.1171875
TUCE     .1049512224E-01   .19482854E-01   .539      .5901      21.937500
PSI      .3785547879      .13917274      2.720      .0065      .43750000

```

Normal exit from iterations. Exit status=0.

```

+-----+
| Binomial Probit Model
| Maximum Likelihood Estimates
| Dependent variable          GRADE
| Weighting variable          ONE
| Number of observations      32
| Iterations completed        6
| Log likelihood function     -12.81880
| Restricted log likelihood    -20.59173
| Chi-squared                 15.54585
| Degrees of freedom          3
| Significance level           .1404896E-02
+-----+

```

```

+-----+-----+-----+-----+-----+-----+
|Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
+-----+-----+-----+-----+-----+-----+
Index function for probability
Constant -7.452319647      2.5424723      -2.931      .0034
GPA      1.625810039      .69388249      2.343      .0191      3.1171875
TUCE     .5172894549E-01   .83890261E-01   .617      .5375      21.937500
PSI      1.426332342      .59503790      2.397      .0165      .43750000

```

Matrix: Las
[4,4]

```

+-----+
| Partial derivatives of E[y] = F[*] with
| respect to the vector of characteristics.
| They are computed at the means of the Xs.
| Observations used for means are All Obs.
+-----+

```

```

+-----+-----+-----+-----+-----+-----+
|Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
+-----+-----+-----+-----+-----+-----+
Index function for probability
Constant -2.444733653      .75885194      -3.222      .0013
GPA      .5333470255      .23246407      2.294      .0218      3.1171875
TUCE     .1696968191E-01   .27119788E-01   .626      .5315      21.937500
PSI      .4679083617      .18764238      2.494      .0126      .43750000

```

Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability.

```

-----+-----+-----+-----+
Actual  Predicted
-----+-----+-----+-----+
        0      1      Total
-----+-----+-----+-----+
0       18      3       21
1        3      8       11
-----+-----+-----+-----+
Total   21     11       32

```

```

--> Calc      ; List ; Scale = N01(b'xbar) $
SCALE        = .32805002591068580D+00
--> Matrix    ; List ; ME = Scale * b $

```

Matrix ME has 4 rows and 1 columns.

```

+-----+
1 | -.2444734D+01
2 | .5333470D+00
3 | .1696968D-01
4 | .4679084D+00
--> Logit      ; Lhs = Grade ; Rhs = X ; marginal Effects $

```

```

+-----+
| Multinomial logit model
| There are 2 outcomes for LH variable GRADE
| These are the OLS start values based on the
| binary variables for each outcome Y(i) = j.
| Coefficients for LHS=0 outcome are set to 0.0
+-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Characteristics in numerator of Prob[Y = 1]					
Constant	-1.498017120	.52388862	-2.859	.0042	
GPA	.4638516793	.16195635	2.864	.0042	3.1171875
TUCE	.1049512224E-01	.19482854E-01	.539	.5901	21.937500
PSI	.3785547879	.13917274	2.720	.0065	.43750000

Normal exit from iterations. Exit status=0.

Multinomial Logit Model	
Maximum Likelihood Estimates	
Dependent variable	GRADE
Weighting variable	ONE
Number of observations	32
Iterations completed	6
Log likelihood function	-12.88963
Restricted log likelihood	-20.59173
Chi-squared	15.40419
Degrees of freedom	3
Significance level	.1501878E-02

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Characteristics in numerator of Prob[Y = 1]					
Constant	-13.02134648	4.9313241	-2.641	.0083	
GPA	2.826112525	1.2629411	2.238	.0252	3.1171875
TUCE	.9515765670E-01	.14155420	.672	.5014	21.937500
PSI	2.378687596	1.0645642	2.234	.0255	.43750000

Matrix: Las
[4,4]

Partial derivatives of probabilities with respect to the vector of characteristics. They are computed at the means of the Xs. Observations used for means are All Obs.

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Marginal effects on Prob[Y = 1]					
Constant	-2.459760743	.81771031	-3.008	.0026	
GPA	.5338588183	.23703797	2.252	.0243	3.1171875
TUCE	.1797548884E-01	.26236909E-01	.685	.4933	21.937500
PSI	.4493392735	.19676264	2.284	.0224	.43750000

Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability.

Actual	Predicted		Total
	0	1	
0	18	3	21
1	3	8	11
Total	21	11	32

```
--> Calc ; List ; Scale = LGD(b'xbar) $
SCALE = .18890218048721940D+00
--> Matrix ; List ; ME = Scale * b $

Matrix ME has 4 rows and 1 columns.
1
+-----+
1 | -.2459761D+01
2 | .5338588D+00
3 | .1797549D-01
4 | .4493393D+00
--> Create ; d0 = 1-Grade ; d1 = Grade $
--> Maximize ; Fcn = -d0*exp(b1'x) + d1*log(1-exp(-exp(b1'x)))
; Labels = b1,b2,b3,b4
; Start = b $
```

Normal exit from iterations. Exit status=0.

User Defined Optimization	
Maximum Likelihood Estimates	
Dependent variable	Function
Weighting variable	ONE
Number of observations	32

```

      | Iterations completed          10 |
      | Log likelihood function      -13.00800 |
      +-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
B1	-10.03142629	3.4448183	-2.912	.0036	
B2	2.293553481	.99127580	2.314	.0207	
B3	.4115615961E-01	.12327466	.334	.7385	
B4	1.562276316	.85886363	1.819	.0689	

Matrix: Las
[4,4]

```

--> Calc ; List ; Scale = exp(-exp(b'xbar))*exp(b'xbar) $
SCALE = .20817920176287030D+00
--> Matrix ; List ; ME = Scale * b $

```

Matrix ME has 4 rows and 1 columns.

```

1 |
2 |
3 |
4 |

```

```

--> /*
-----+-----
Ordinary least squares regression Weighting variable = none
Dep. var. = GRADE Mean= .3437500000 , S.D.= .4825587044
Model size: Observations = 32, Parameters = 4, Deg.Fr.= 28
-----+-----

```

Variable	Coefficient	Standard Error	t-ratio	P[T >t]	Mean of X
Constant	-1.498017120	.52388862	-2.859	.0079	
GPA	.4638516793	.16195635	2.864	.0078	3.1171875
TUCE	.1049512224E-01	.19482854E-01	.539	.5944	21.937500
PSI	.3785547879	.13917274	2.720	.0111	.43750000

```

SCALE = .100000000000000000D+01
Matrix ME has 4 rows and 1 columns.

```

```

1 |
2 |
3 |
4 |

```

```

-----+-----
Binomial Probit Model
Number of observations          32
Iterations completed            6
Log likelihood function        -12.81880
Restricted log likelihood      -20.59173
Chi-squared                    15.54585
Degrees of freedom              3
Significance level              .1404896E-02
-----+-----

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-7.452319647	2.5424723	-2.931	.0034	
GPA	1.625810039	.69388249	2.343	.0191	3.1171875
TUCE	.5172894549E-01	.83890261E-01	.617	.5375	21.937500
PSI	1.426332342	.59503790	2.397	.0165	.43750000

```

-----+-----
Partial derivatives of E[y] = F[*] with
respect to the vector of characteristics.
They are computed at the means of the Xs.
-----+-----

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
----------	-------------	----------------	----------	----------	-----------

Index f
Command is more than 2500 characters.

```

--> Constant -2.444733653 .75885194 -3.222 .0013
GPA .5333470255 .23246407 2.294 .0218 3.1171875
TUCE .1696968191E-01 .27119788E-01 .626 .5315 21.937500
PSI .4679083617 .18764238 2.494 .0126 .43750000
SCALE = .32805002591068580D+00
Matrix ME has 4 rows and 1 columns.

```

```

1 |
2 |
3 |
4 |

```

```

Multinomial Logit Model
Dependent variable          GRADE
Number of observations      32
Iterations completed       6
Log likelihood function    -12.88963
Restricted log likelihood   -20.59173
Chi-squared                15.40419
Degrees of freedom         3
Significance level         .1501878E-02
-----+-----
|Variable| Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X|
-----+-----
Characteristics in numerator of Prob[Y = 1]
Constant -13.02134648      4.9313241      -2.641      .0083
GPA       2.826112525       1.2629411       2.238      .0252      3.1171875
TUCE     .9515765670E-01      .14155420       .672      .5014      21.9375000
PSI      2.378687596       1.0645642       2.234      .0255      .437500000
-----+-----
|Partial derivatives of probabilities with
|respect to the vector of characteristics.
|They are computed at the means of the Xs.
|Observations used for means are All Obs.
|-----+-----
|Variable| Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X|
-----+-----
Marginal effects on Prob[Y = 1]
Constant -2.459760743      .81771031      -3.008      .0026
GPA       .5338588183             .23703797       2.252      .0243      3.1171875
TUCE     .1797548884E-01      .26236909E-01   .685      .4933      21.9375000
PSI      .4493392735             .19676264       2.284      .0224      .437500000
SCALE = .18890218048721940D+00
Matrix ME has 4 rows and 1 columns.
-----+-----
1| -.2459761D+01
2| .5338588D+00
3| .1797549D-01
4| .4493393D+00
-----+-----
Command is more than 2500 characters.
--> User Defined Optimization
Number of observations      32
Iterations completed       10
Log likelihood function    -13.00800
-----+-----
|Variable| Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X|
-----+-----
B1      -10.03142629         3.4448183      -2.912      .0036
B2       2.293553481         .99127580       2.314      .0207
B3       4.115615961E-01     .12327466       .334      .7385
B4       1.562276316          .85886363       1.819      .0689
SCALE = .20817920176287030D+00
Matrix ME has 4 rows and 1 columns.
-----+-----
1| -.2088334D+01
2| .4774701D+00
3| .8567856D-02
4| .3252334D+00
*/
--> Probit ; Lhs = Grade ; Rhs = X $
-----+-----
Dependent variable is binary, y=0 or y not equal 0
Ordinary least squares regression Weighting variable = none
Dep. var. = GRADE Mean= .3437500000 , S.D.= .4825587044
Model size: Observations = 32, Parameters = 4, Deg.Fr.= 28
Residuals: Sum of squares= 4.216473951 , Std.Dev.= .38806
Fit: R-squared= .415900, Adjusted R-squared = .35332
Model test: F[ 3, 28] = 6.65, Prob value = .00157
Diagnostic: Log-L = -12.9782, Restricted(b=0) Log-L = -21.5812
LogAmemiyaPrCrt.= -1.775, Akaike Info. Crt.= 1.061
-----+-----
|Variable| Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X|
-----+-----
Constant -1.498017120         .52388862      -2.859      .0042
GPA       .4638516793             .16195635       2.864      .0042      3.1171875
TUCE     .1049512224E-01      .19482854E-01   .539      .5901      21.9375000
PSI      .3785547879             .13917274       2.720      .0065      .437500000
-----+-----
Normal exit from iterations. Exit status=0.
-----+-----
|Binomial Probit Model
|Maximum Likelihood Estimates
|-----+-----

```

```

Dependent variable          GRADE
Weighting variable         ONE
Number of observations      32
Iterations completed       6
Log likelihood function    -12.81880
Restricted log likelihood  -20.59173
Chi-squared               15.54585
Degrees of freedom        3
Significance level        .1404896E-02

```

```

-----+-----+-----+-----+-----+-----+
|Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
-----+-----+-----+-----+-----+-----+
Index function for probability
Constant -7.452319647      2.5424723      -2.931      .0034
GPA      1.625810039      .69388249     2.343      .0191      3.1171875
TUCE     .5172894549E-01   .83890261E-01 .617      .5375     21.937500
PSI      1.426332342      .59503790     2.397      .0165     .43750000

```

Matrix: Lhs
[4,4]

Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability.

```

-----+-----+-----+-----+
Predicted
Actual  0  1  | Total
-----+-----+-----+
0       18  3  | 21
1        3  8  | 11
-----+-----+-----+
Total   21  11 | 32

```

```

--> Sample ; 1 - 100 $
--> Create ; GPAi = Trn(2,.02)
; YesPSI = PHI(b(1) + b(2)*GPAi + b(3)*xbr(Tuce) + b(4))
; NOPSI  = PHI(b(1) + b(2)*GPAi + b(3)*xbr(Tuce) ) $
--> Calc ; MeanYes = PHI(b(1) + b(2)*xbr(GPA) + b(3)*xbr(Tuce) + b(4))
; MeanNo  = PHI(b(1) + b(2)*xbr(GPA) + b(3)*xbr(Tuce) )
; MeanGPA = Xbr(GPA) $
--> Plot ; Lhs = GPAi
; Rhs = YESPSI, NOPSI
; Bars= MEANYes, MEANNo
; Spikes = MeanGPA
; Limits = 0,1 ; EndPoints = 2,4 ; Fill ; Yaxis=P[Grd=1]
; Title=Effect of PSI on Probabilities $
The specification ;YAX is not recognized.
--> Sample ; 1 - 32 $
--> Probit ; Lhs = Grade ; Rhs = X $

```

```

-----+-----+-----+-----+
Dependent variable is binary, y=0 or y not equal 0
Ordinary least squares regression Weighting variable = none
Dep. var. = GRADE Mean= .3437500000 , S.D.= .4825587044
Model size: Observations = 32, Parameters = 4, Deg.Fr.= 28
Residuals: Sum of squares= 4.216473951 , Std.Dev.= .38806
Fit: R-squared= .415900, Adjusted R-squared = .35332
Model test: F[ 3, 28] = 6.65, Prob value = .00157
Diagnostic: Log-L = -12.9782, Restricted(b=0) Log-L = -21.5812
LogAmemiyaPrCrt.= -1.775, Akaike Info. Crt.= 1.061

```

```

-----+-----+-----+-----+
|Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
-----+-----+-----+-----+-----+
Constant -1.498017120      .52388862     -2.859     .0042
GPA      .4638516793      .16195635     2.864     .0042     3.1171875
TUCE     .1049512224E-01   .19482854E-01 .539     .5901     21.937500
PSI      .3785547879      .13917274     2.720     .0065     .43750000

```

Normal exit from iterations. Exit status=0.

```

-----+-----+-----+-----+
Binomial Probit Model
Maximum Likelihood Estimates
Dependent variable          GRADE
Weighting variable         ONE
Number of observations      32
Iterations completed       6
Log likelihood function    -12.81880
Restricted log likelihood  -20.59173
Chi-squared               15.54585
Degrees of freedom        3
Significance level        .1404896E-02

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Index function for probability					
Constant	-7.452319647	2.5424723	-2.931	.0034	
GPA	1.625810039	.69388249	2.343	.0191	3.1171875
TUCE	.5172894549E-01	.83890261E-01	.617	.5375	21.937500
PSI	1.426332342	.59503790	2.397	.0165	.43750000

Matrix: Las
[4,4]

Frequencies of actual & predicted outcomes
 Predicted outcome has maximum probability.

Actual	Predicted		Total
	0	1	
0	18	3	21
1	3	8	11
Total	21	11	32

```
--> Sample ; 1 - 100 $
--> Create ; GPai = Trn(2,.02)
; YesPSI = PHI(b(1) + b(2)*GPai + b(3)*xbr(Tuce) + b(4))
; NOPSI = PHI(b(1) + b(2)*GPai + b(3)*xbr(Tuce) ) $
--> Calc ; MeanYes = PHI(b(1) + b(2)*xbr(GPA) + b(3)*xbr(Tuce) + b(4))
; MeanNo = PHI(b(1) + b(2)*xbr(GPA) + b(3)*xbr(Tuce) )
; MeanGPA = Xbr(GPA) $
--> Plot ; Lhs = GPai
; Rhs = YESPSI, NOPSI
; Bars= MEANYes,MEANNo
; Spikes = MeanGPA
; Limits = 0,1 ; EndPoints = 2,4 ; Fill ; Yaxis=P[Grade=1]
; Title=Effect of PSI on Probabilities $
The specification ;YAX is not recognized.
--> Sample ; 1 - 32 $
--> Probit ; Lhs = Grade ; Rhs = X $
```

Dependent variable is binary, y=0 or y not equal 0	
Ordinary least squares regression	Weighting variable = none
Dep. var. = GRADE	Mean= .3437500000 , S.D.= .4825587044
Model size: Observations = 32, Parameters = 4, Deg.Fr.= 28	
Residuals: Sum of squares= 4.216473951	Std.Dev.= .38806
Fit: R-squared= .415900, Adjusted R-squared = .35332	
Model test: F[3, 28] = 6.65, Prob value = .00157	
Diagnostic: Log-L = -12.9782, Restricted(b=0) Log-L = -21.5812	
LogAmemiyaPrCrt.= -1.775, Akaike Info. Crt.= 1.061	

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-1.498017120	.52388862	-2.859	.0042	
GPA	.4638516793	.16195635	2.864	.0042	3.1171875
TUCE	.1049512224E-01	.19482854E-01	.539	.5901	21.937500
PSI	.3785547879	.13917274	2.720	.0065	.43750000

Normal exit from iterations. Exit status=0.

Binomial Probit Model	
Maximum Likelihood Estimates	
Dependent variable	GRADE
Weighting variable	ONE
Number of observations	32
Iterations completed	6
Log likelihood function	-12.81880
Restricted log likelihood	-20.59173
Chi-squared	15.54585
Degrees of freedom	3
Significance level	.1404896E-02

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Index function for probability					
Constant	-7.452319647	2.5424723	-2.931	.0034	
GPA	1.625810039	.69388249	2.343	.0191	3.1171875
TUCE	.5172894549E-01	.83890261E-01	.617	.5375	21.937500
PSI	1.426332342	.59503790	2.397	.0165	.43750000

Matrix: Las
[4,4]

Frequencies of actual & predicted outcomes
 Predicted outcome has maximum probability.

Actual	Predicted		Total
	0	1	
0	18	3	21
1	3	8	11
Total	21	11	32

```
--> Sample ; 1 - 100 $
--> Create ; GPai = Trn(2,.02)
; YesPSI = PHI(b(1) + b(2)*GPai + b(3)*xbr(Tuce) + b(4))
; NOPSI = PHI(b(1) + b(2)*GPai + b(3)*xbr(Tuce) ) $
--> Calc ; MeanYes = PHI(b(1) + b(2)*xbr(GPA) + b(3)*xbr(Tuce) + b(4))
; MeanNo = PHI(b(1) + b(2)*xbr(GPA) + b(3)*xbr(Tuce) )
; MeanGPA = Xbr(GPA) $
--> Plot ; Lhs = GPai
; Rhs = YESPSI, NOPSI
; Bars= MEANYes,MEANNo
; Spikes = MeanGPA
; Limits = 0,1 ; EndPoints = 2,4 ; Fill ; Yaxis=P[Grade=1]
; Title=Effect of PSI on Probabilities $
The specification ;YAX is not recognized.
--> Sample ; 1 - 32 $
--> Probit ; Lhs = Grade ; Rhs = X $
```

```
-----+-----+-----+-----+-----+-----+
Dependent variable is binary, y=0 or y not equal 0
Ordinary least squares regression Weighting variable = none
Dep. var. = GRADE Mean= .343750000 , S.D.= .4825587044
Model size: Observations = 32, Parameters = 4, Deg.Fr.= 28
Residuals: Sum of squares= 4.216473951 , Std.Dev.= .38806
Fit: R-squared= .415900, Adjusted R-squared = .35332
Model test: F[ 3, 28] = 6.65, Prob value = .00157
Diagnostic: Log-L = -12.9782, Restricted(b=0) Log-L = -21.5812
LogAmemiyaPrCrt.= -1.775, Akaike Info. Crt.= 1.061
-----+-----+-----+-----+-----+-----+
|Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
-----+-----+-----+-----+-----+-----+
Constant -1.498017120 .52388862 -2.859 .0042
GPA .4638516793 .16195635 2.864 .0042 3.1171875
TUCE .1049512224E-01 .19482854E-01 .539 .5901 21.937500
PSI .3785547879 .13917274 2.720 .0065 .43750000
-----+-----+-----+-----+-----+-----+
Normal exit from iterations. Exit status=0.
```

```
-----+-----+-----+-----+-----+-----+
Binomial Probit Model
Maximum Likelihood Estimates
Dependent variable GRADE
Weighting variable ONE
Number of observations 32
Iterations completed 6
Log likelihood function -12.81880
Restricted log likelihood -20.59173
Chi-squared 15.54585
Degrees of freedom 3
Significance level .1404896E-02
-----+-----+-----+-----+-----+-----+
|Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
-----+-----+-----+-----+-----+-----+
Index function for probability
Constant -7.452319647 2.5424723 -2.931 .0034
GPA 1.625810039 .69388249 2.343 .0191 3.1171875
TUCE .5172894549E-01 .83890261E-01 .617 .5375 21.937500
PSI 1.426332342 .59503790 2.397 .0165 .43750000
-----+-----+-----+-----+-----+-----+

```

Matrix: Las
[4,4]

Frequencies of actual & predicted outcomes
 Predicted outcome has maximum probability.

Actual	Predicted		Total
	0	1	
0	18	3	21
1	3	8	11
Total	21	11	32

```

-----+-----
0      18  3  |  21
1      3  8  |  11
-----+-----
Total  21  11 |  32
--> Sample ; 1 - 100 $
--> Create ; GPAi = Trn(2,.02)
; YesPSI = PHI(b(1) + b(2)*GPAi + b(3)*xbr(Tuce) + b(4))
; NOPSI  = PHI(b(1) + b(2)*GPAi + b(3)*xbr(Tuce) ) $
--> Calc ; MeanYes = PHI(b(1) + b(2)*xbr(GPA) + b(3)*xbr(Tuce) + b(4))
; MeanNo  = PHI(b(1) + b(2)*xbr(GPA) + b(3)*xbr(Tuce) )
; MeanGPA = Xbr(GPA) $
--> Plot ; Lhs = GPAi
; Rhs = YESPSI, NOPSI
; Bars= MEANYes,MEANNo
; Spikes = MeanGPA
; Limits = 0,1 ; EndPoints = 2,4 ; Fill ; Yaxis = P[Grade=1]
; Title=Effect of PSI on Probabilities $
The specification ;YAX is not recognized.
--> Sample ; 1 - 32 $
--> Probit ; Lhs = Grade ; Rhs = X $

```

```

-----+-----
Dependent variable is binary, y=0 or y not equal 0
Ordinary least squares regression Weighting variable = none
Dep. var. = GRADE Mean= .343750000 , S.D.= .4825587044
Model size: Observations = 32, Parameters = 4, Deg.Fr.= 28
Residuals: Sum of squares= 4.216473951 , Std.Dev.= .38806
Fit: R-squared= .415900, Adjusted R-squared = .35332
Model test: F[ 3, 28] = 6.65, Prob value = .00157
Diagnostic: Log-L = -12.9782, Restricted(b=0) Log-L = -21.5812
LogAmemiyaPrCrt.= -1.775, Akaike Info. Crt.= 1.061
-----+-----

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-1.498017120	.52388862	-2.859	.0042	
GPA	.4638516793	.16195635	2.864	.0042	3.1171875
TUCE	.1049512224E-01	.19482854E-01	.539	.5901	21.937500
PSI	.3785547879	.13917274	2.720	.0065	.43750000

Normal exit from iterations. Exit status=0.

```

-----+-----
Binomial Probit Model
Maximum Likelihood Estimates
Dependent variable          GRADE
Weighting variable         ONE
Number of observations      32
Iterations completed        6
Log likelihood function     -12.81880
Restricted log likelihood    -20.59173
Chi-squared                 15.54585
Degrees of freedom          3
Significance level          .1404896E-02
-----+-----

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Index function for probability					
Constant	-7.452319647	2.5424723	-2.931	.0034	
GPA	1.625810039	.69388249	2.343	.0191	3.1171875
TUCE	.5172894549E-01	.83890261E-01	.617	.5375	21.937500
PSI	1.426332342	.59503790	2.397	.0165	.43750000

Matrix: Las
[4,4]

Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability.

Actual	Predicted		Total
	0	1	
0	18	3	21
1	3	8	11
Total	21	11	32

```

--> Sample ; 1 - 100 $
--> Create ; GPAi = Trn(2,.02)
; YesPSI = PHI(b(1) + b(2)*GPAi + b(3)*xbr(Tuce) + b(4))
; NOPSI  = PHI(b(1) + b(2)*GPAi + b(3)*xbr(Tuce) ) $
--> Calc ; MeanYes = PHI(b(1) + b(2)*xbr(GPA) + b(3)*xbr(Tuce) + b(4))
; MeanNo  = PHI(b(1) + b(2)*xbr(GPA) + b(3)*xbr(Tuce) )

```

```

; MeanGPA = Xbr(GPA) $
--> Plot ; Lhs = GPai
; Rhs = YESPSI, NOPSI
; Bars= MEANYes,MEANNo
; Spikes = MeanGPA
; Limits = 0,1 ; EndPoints = 2,4 ; Fill ;
; Title=Effect of PSI on Probabilities $
--> Sample ; 1 - 32 $

```

Shazam Program for computing Tobit Model

```

*          TOBIT REGRESSION
*
* Data from Judge, Hill, Griffiths, Lutkepohl and Lee,
* INTRODUCTION TO THE THEORY AND PRACTICE OF ECONOMETRICS,
* Second Edition, Wiley, 1988, Table 19.2, page 800.
SAMPLE 1 20
READ (judge19.txt) Y X
TOBIT Y X / LIST
TEST X/Y
TEST CONSTANT/Y
STOP

```

0.	1
0.	2
0.	3
0.	4
3.134800	5
3.508000	6
0.8312000	7
8.006400	8
0.	9
0.	10
2.935200	11
3.904800	12
6.514400	13
5.977200	14
3.726000	15
10.41240	16
16.90640	17
9.296800	18
7.891600	19
14.21640	20

```

|_*          TOBIT REGRESSION
|_*
|_* Data from Judge, Hill, Griffiths, Lutkepohl and Lee,
|_* INTRODUCTION TO THE THEORY AND PRACTICE OF ECONOMETRICS,
|_* Second Edition, Wiley, 1988, Table 19.2, page 800.
|_*SAMPLE 1 20
|_*READ (judge19.txt) Y X
UNIT 88 IS NOW ASSIGNED TO: judge19.txt
    2 VARIABLES AND          20 OBSERVATIONS STARTING AT OBS      1

```

|_*TOBIT Y X / LIST

REQUIRED MEMORY IS PAR= 1 CURRENT PAR= 2000
FOR MAXIMUM EFFICIENCY USE AT LEAST PAR= 2

```

TOBIT ANALYSIS, LIMIT= 0.00 25 MAX ITERATIONS
6 LIMIT OBSERVATIONS
14 NON-LIMIT OBSERVATIONS

ITERATION 0 NORMALIZED COEFFICIENTS
0.24935 -1.5481 0.27952

ITERATION 1 NORMALIZED COEFFICIENTS
0.24816 -1.5783 0.27538

ITERATION 2 NORMALIZED COEFFICIENTS
0.24820 -1.5786 0.27541

ITERATION 3 NORMALIZED COEFFICIENTS
0.24820 -1.5786 0.27541

FIRST DERIVATIVES OF LOG OF LIKELIHOOD FUNCTION EVALUATED AT MAXIMUM
0.28421709E-13 0.13322676E-14 -0.18651747E-13

NUMBER OF ITERATIONS = 3

DEPENDENT VARIABLE = Y
VARIANCE OF THE ESTIMATE = 13.184
STANDARD ERROR OF THE ESTIMATE = 3.6310

ASYMPTOTIC
VARIABLE NORMALIZED STANDARD T-RATIO REGRESSION ELASTICITY ELASTICITY
COEFFICIENT ERROR COEFFICIENT OF INDEX OF E(Y)
X 0.24820 0.58959E-01 4.2096 0.90120 1.9458 1.9970
CONSTANT -1.5786 0.60413 -2.6130 -5.7319
Y 0.27541 0.53292E-01 5.1679

THE PREDICTED PROBABILITY OF Y > LIMIT GIVEN AVERAGE X(I) = 0.8479
THE OBSERVED FREQUENCY OF Y > LIMIT IS = 0.7000
AT MEAN VALUES OF ALL X(I), E(Y) = 4.0177

DEPENDENT VARIABLE
OB INDEX PROB(X) DENSITY(X) OBSERVED EXPECTED CONDITIONAL
1 -1.3304 0.91692E-01 0.16465 0.0000 0.15491 -----
2 -1.0822 0.13958 0.22212 0.0000 0.25805 -----
3 -0.83402 0.20214 0.28175 0.0000 0.41091 -----
4 -0.58582 0.27900 0.33604 0.0000 0.62670 -----
5 -0.33762 0.36782 0.37684 3.1348 0.91739 2.4941
6 -0.89429E-01 0.46437 0.39735 3.5080 1.2920 2.7822
7 0.15877 0.56307 0.39395 0.83120 1.7550 3.1169
8 0.40696 0.65798 0.36724 8.0064 2.3057 3.5042
9 0.65516 0.74382 0.32189 0.0000 2.9382 -----
10 0.90335 0.81683 0.26528 0.0000 3.6425 -----
11 1.1515 0.87525 0.20557 2.9352 4.4061 5.0341
12 1.3997 0.91921 0.14978 3.9048 5.2157 5.6741
13 1.6479 0.95032 0.10261 6.5144 6.0590 6.3758
14 1.8961 0.97103 0.66099E-01 5.9772 6.9254 7.1321
15 2.1443 0.98400 0.40034E-01 3.7260 7.8069 7.9338
16 2.3925 0.99163 0.22799E-01 10.412 8.6974 8.7708
17 2.6407 0.99586 0.12208E-01 16.906 9.5932 9.6330
18 2.8889 0.99807 0.61466E-02 9.2968 10.492 10.512
19 3.1371 0.99915 0.29098E-02 7.8916 11.392 11.401
20 3.3853 0.99964 0.12952E-02 14.216 12.292 12.297

LOG-LIKELIHOOD FUNCTION= -41.255240
MEAN-SQUARE ERROR= 8.0589568
MEAN ERROR=-0.40295353E-02
MEAN ABSOLUTE ERROR= 2.1501402
SQUARED CORRELATION BETWEEN OBSERVED AND EXPECTED VALUES= 0.66061
|_TEST X/Y
TEST VALUE = 0.90120 STD. ERROR OF TEST VALUE 0.17186
ASYMPTOTIC NORMAL STATISTIC = 5.2437210 P-VALUE= 0.00000
WALD CHI-SQUARE STATISTIC = 27.496610 WITH 1 D.F. P-VALUE= 0.00000
UPPER BOUND ON P-VALUE BY CHEBYCHEV INEQUALITY = 0.03637
|_TEST CONSTANT/Y
TEST VALUE = -5.7319 STD. ERROR OF TEST VALUE 2.2383
ASYMPTOTIC NORMAL STATISTIC = -2.5608774 P-VALUE= 0.01044
WALD CHI-SQUARE STATISTIC = 6.5580928 WITH 1 D.F. P-VALUE= 0.01044
UPPER BOUND ON P-VALUE BY CHEBYCHEV INEQUALITY = 0.15248
|_STOP
TYPE COMMAND

```