# TEACHING MICROECONOMETRICS USING STATA

1. Models with discrete dependent variables

2. Censored dependent variables

# Models with discrete dependent variables

Can have qualitative response models where the dependent variable is **discrete** rather than a **continuous** variable. Types of discrete choice models:

*a) Dichotomous, binary or dummy variables*

Such models take on the value of zero or one. For example modelling the probability of being unemployed:

$$y_i = \begin{cases} 1 & Unemployed \\ 0 & Employed \end{cases}$$

or the probability of being in debt:

$$y_i = \begin{cases} 1 & Debtor \\ 0 & NonDebtor \end{cases}$$

*b) Polychotomous variables*

These take on a discrete number and can be split into:

*i. unordered variables*

These are variables for which there is no natural ranking of the alternatives. For example for a sample of commuters we might want to construct a variable:

$$y_i = \begin{cases} 0 \;\; if \;\; person\, i\, not\, in\, labour\, market \\ 1 \;\; if \;\; person\, i\, is\, an\, employee \\ 2 \;\; if \;\; person\, i\, is\, self\;\, employed \\ 3 \;\; if \;\; person\, i\, is\, unemployed \end{cases}$$

*ii. Ordered variables*

With such variables the outcomes have a natural ranking. For example suppose we have a sample on individual's health status:

$$y_i = \begin{cases} 0 & if \ \ person\, i\, is\, in\ poor\, health \\ 1 & if \ \ person\, i\, is\, in\ fair\, health \\ 2 & if \ \ person\, i\, is\, in\, excellent\, health \end{cases}$$

Another example of an ordered variable is a *sequential variable*:

$$y_i = \begin{cases} 0 & O'levels \\ 1 & A'levels \\ 2 & Graduate\, \mathrm{deg}ree \\ 3 & Postgraduate\, \mathrm{deg}ree \end{cases}$$

## I. <u>Ordered Choice Models</u>

### The Ordered Probit Model

The model is built around the latent variable framework in the same way as the binomial probit model:

$$y^* = \boldsymbol{x}'\boldsymbol{\beta} + \varepsilon$$

where $y^*$ is unobserved. What we do observe is

$$\begin{array}{lll}
y = 0 & \text{if} & y^* \leq 0 \\
y = 1 & \text{if} & 0 \leq y^* \leq \mu_1 \\
y = 2 & \text{if} & \mu_1 \leq y^* \leq \mu_2 \\
\vdots & \vdots & \vdots \\
y = J & \text{if} & \mu_{J-1} \leq y^*
\end{array}$$

This adheres to a type of censoring.

The $\mu$'s are unknown parameters to be estimated along with the $\boldsymbol{\beta}$.

Basing the above upon having normally distributed errors across observations, normalising the mean and variance to 0 and 1 respectively (as with the binomial probit), we have the following probabilities:

$$
\begin{array}{lcl}
prob\left(y=0|\boldsymbol{x}\right) & = & \Phi\left(-\boldsymbol{x}'\boldsymbol{\beta}\right) \\
prob\left(y=1\,|\boldsymbol{x}\right) & = & \Phi\left(\mu_1-\boldsymbol{x}'\boldsymbol{\beta}\right)-\Phi\left(-\boldsymbol{x}'\boldsymbol{\beta}\right) \\
prob\left(y=2|\boldsymbol{x}\right) & = & \Phi\left(\mu_2-\boldsymbol{x}'\boldsymbol{\beta}\right)-\Phi\left(\mu_1-\boldsymbol{x}'\boldsymbol{\beta}\right) \\
\vdots & \vdots & \vdots \\
prob\left(y=J\,|\boldsymbol{x}\right) & = & 1-\Phi\left(\mu_{j-1}-\boldsymbol{x}'\boldsymbol{\beta}\right)
\end{array}
$$

For the probabilities to be positive we must have $0<\mu_1<\mu_2<\cdots<\mu_{J-1}$

**EXAMPLE**

- The 1998 *Workplace Employee Relations Survey WERS* (Department of Trade and Industry, 1999) can be used to model EFFORT.

- 1998 *WERS* has matched employer-employee information and is a nationally representative survey of workplaces with 10+ employees.

- The survey offers comprehensive information on a sample of 28,215 employees working in 1,782 establishments though our final data set (due to missing values) comprises 19,510 employees from 1,753 workplaces.

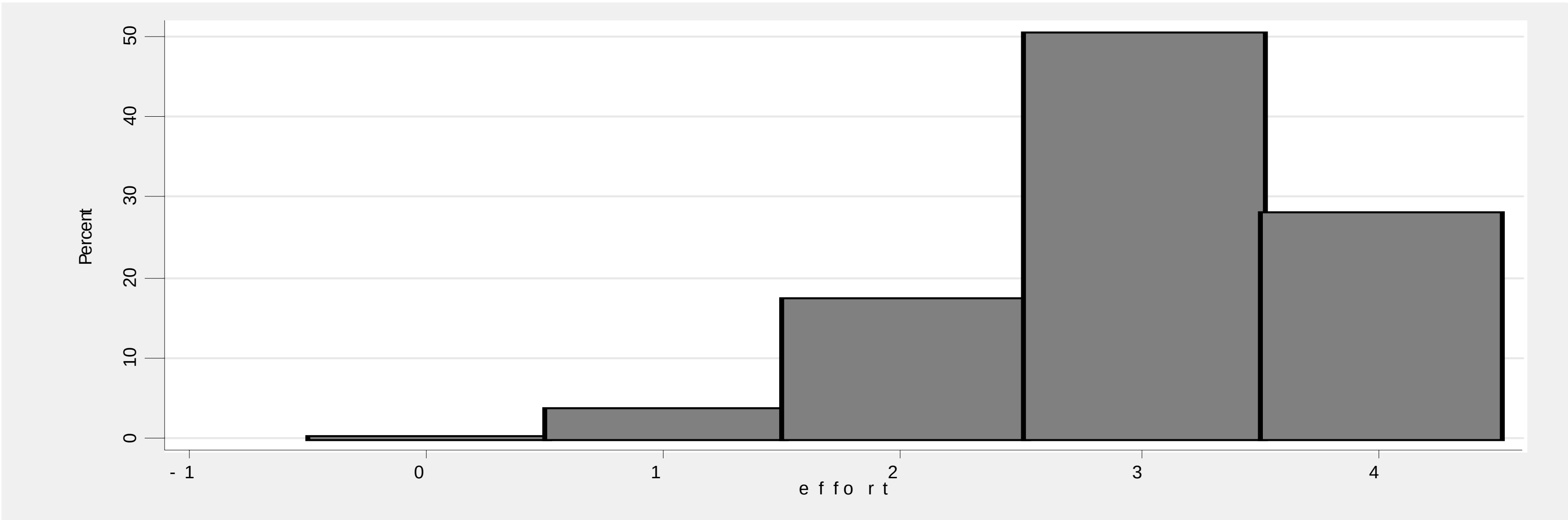
A question asked to employees is:

*Do you agree or disagree that your job requires that you work very hard?*

The responses are categorized as:

$$Effort = \begin{cases} 4 = strongly\ agree \\ 3 = agree \\ 2 = neither\ agree\ nor\ disagree \\ 1 = disagree \\ 0 = strongly\ disagree \end{cases}$$

*Histogram of Effort*




**Clearly this variable has a natural ranking.**

Model effort as:

$$Effort_{fi}^{*} = \boldsymbol{x}_{fi}'\boldsymbol{\beta} + \varepsilon_{fi}$$

where $Effort^{*}$ is the unobserved propensity of an individual $i$ employed in firm $f$ to exert effort, a latent variable; $Effort$ is the individual's observed level of effort.

Variables used to explain effort are the relative wage on offer in the firm, age, gender, ethnicity, health status, contract type, union membership and firm size.

| **VARIABLE** | **DESCRIPTION** |
|---|---|
| *Effort (Eff1)* | *Do you agree or disagree that your job requires that you work very hard?*<br>Index 0=strongly disagree, 1=disagree, 2=neither agree nor disagree, 3=agree, 4=strongly agree |
| *Relwfirm* | Log individuals wage relative to the average firm wage |
| *Male* | Dummy (0/1) equals 1 if individual is male |
| *White* | Dummy (0/1) equals 1 if individual is white |
| *Health* | Dummy (0/1) equals 1 if the individual is in good health |
| *Perm* | Dummy (0/1) equals 1 if the individual has a permanent contract |
| *Tumem* | Dummy (0/1) equals 1 if the individual is a trade union member |
| *Emp* | Number of employees in the firm where the individual works |
| *Empsq* | Number of employees squared |

$$Effort_{fi} = \beta_0 + \beta_1 Relwfirm_{fi} + \beta_2 Male_{fi} + \beta_3 White_{fi} + \beta_4 Health_{fi} + \beta_5 Perm_{fi}$$

$$+ \beta_6 Tumem_{fi} + \beta_7 Emp_{fi} + \beta_7 Empsq_{fi} + \varepsilon_{fi}$$

```
#delimit;
clear;
set mem 100m;
set mat 100;
set more off;
use "E:\karl's files\stata\L5-6.dta", clear;
oprob eff1 relwfirm male white health perm tumem emp empsq;
predict pp0 pp1 pp2 pp3 pp4, p;
mfx compute, predict(outcome(1));
mfx compute, predict(outcome(2));
mfx compute, predict(outcome(3));
mfx compute, predict(outcome(4));
```

## RESULTS

```
oprob eff1 relwfirm male white health perm  tumem emp empsq;

Ordered probit estimates                          Number of obs   =      19510
                                                  LR chi2(8)      =     470.30
                                                  Prob > chi2     =     0.0000
Log likelihood = -22124.227                       Pseudo R2       =     0.0105
```

| eff1 | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| relwfirm | .2761184 | .0213419 | 12.94 | 0.000 | .2342891 | .3179477 |
| male | -.2547255 | .0160936 | -15.83 | 0.000 | -.2862683 | -.2231826 |
| white | -.1968963 | .0440379 | -4.47 | 0.000 | -.2832089 | -.1105836 |
| health | .0176182 | .017328 | 1.02 | 0.309 | -.016344 | .0515805 |
| perm | .097243 | .0335359 | 2.90 | 0.004 | .0315138 | .1629722 |
| tumem | .1364387 | .0163371 | 8.35 | 0.000 | .1044186 | .1684588 |
| emp | -.0000811 | .0000204 | -3.97 | 0.000 | -.0001211 | -.000041 |
| empsq | 6.11e-09 | 2.26e-09 | 2.70 | 0.007 | 1.67e-09 | 1.06e-08 |
| _cut1 | -3.063746 | .0722249 | (Ancillary parameters) | | | |
| _cut2 | -1.966417 | .0568277 | | | | |
| _cut3 | -1.001023 | .0553364 | | | | |
| _cut4 | .3880693 | .0550301 | | | | |

```
. predict p0 p1 p2 p3 p4, p;
```

# MARGINALS

```
. mfx compute, predict(outcome(1));

Marginal effects after oprobit
      y  = Pr(eff1==1) (predict, outcome(1))
         =    .036615
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
relwfirm |  -.0213413      .00175  -12.18   0.000  -.024776 -.017907  -.071163
    male*|   .0196653      .00136   14.46   0.000      .017   .02233   .513378
   white*|   .0130344      .00249    5.23   0.000   .008152  .017917   .966376
  health*|   -.001369      .00135   -1.01   0.312  -.004023  .001285   .680113
    perm*|  -.0080722      .00299   -2.70   0.007  -.013932 -.002212   .941825
   tumem*|  -.0103698      .00126   -8.25   0.000  -.012834 -.007905    .42081
     emp |   6.27e-06      .00000    3.94   0.000   3.2e-06  9.4e-06   295.934
   empsq |  -4.72e-10      .00000   -2.69   0.007  -8.2e-10 -1.3e-10    509435
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

```
. mfx compute, predict(outcome(2));

Marginal effects after oprobit
      y  = Pr(eff1==2) (predict, outcome(2))
         =    .173017
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
relwfirm |  -.0567923      .00445  -12.75   0.000  -.065522 -.048062  -.071163
    male*|   .0521267      .00334   15.61   0.000   .045581  .058673   .513378
   white*|   .0386936      .00821    4.72   0.000   .022611  .054776   .966376
  health*|  -.0036283      .00357   -1.02   0.310  -.010632  .003375   .680113
    perm*|  -.0203127      .00711   -2.86   0.004  -.034239 -.006386   .941825
   tumem*|  -.0279082      .00334   -8.36   0.000   -.03445 -.021366    .42081
     emp |   .0000167      .00000    3.96   0.000   8.4e-06  .000025   295.934
   empsq |  -1.26e-09      .00000   -2.70   0.007  -2.2e-09 -3.4e-10    509435
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

```
. mfx compute, predict(outcome(3));

Marginal effects after oprobit
      y  = Pr(eff1==3) (predict, outcome(3))
         =  .51022488
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
relwfirm |  -.0126706      .00138   -9.21   0.000  -.015367 -.009974  -.071163
   male*|   .0119798      .00117   10.22   0.000   .009681  .014278   .513378
  white*|   .0166074       .0054    3.08   0.002   .006027  .027188   .966376
 health*|  -.0007844      .00075   -1.05   0.296  -.002255  .000686   .680113
   perm*|  -.0026572      .00045   -5.96   0.000  -.003531 -.001783   .941825
  tumem*|  -.0068794      .00102   -6.77   0.000   -.00887 -.004888    .42081
     emp |   3.72e-06      .00000    3.80   0.000   1.8e-06  5.6e-06   295.934
   empsq |  -2.80e-10      .00000   -2.64   0.008  -4.9e-10 -7.3e-11    509435
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

```
. mfx compute, predict(outcome(4));

Marginal effects after oprobit
      y  = Pr(eff1==4) (predict, outcome(4))
         =  .27804591
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
relwfirm |   .0926318       .00716   12.93   0.000   .078589  .106675  -.071163
    male*|  -.0854741        .0054  -15.84   0.000  -.096053 -.074895   .513378
   white*|  -.0693456       .01618   -4.29   0.000  -.101054 -.037637   .966376
  health*|   .0058994       .00579    1.02   0.308  -.005451   .01725   .680113
    perm*|   .0317714       .01066    2.98   0.003   .010888  .052655   .941825
   tumem*|   .0460378       .00554    8.30   0.000   .035173  .056903    .42081
     emp |  -.0000272       .00001   -3.97   0.000  -.000041 -.000014   295.934
   empsq |   2.05e-09       .00000    2.70   0.007   5.6e-10  3.5e-09    509435
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

## Interpreting the marginal effects

Comparing effort categories 4 to 3 i.e. strongly agreeing to the question:

*Do you agree or disagree that your job requires that you work very hard?*

rather than answering ‘agrees’

Then the impact of the relative wage earned by the individual in comparison to their workmates is that a 1% higher relative wage leads to a 9.3% higher probability of replying in the top category.

The impact of being male leads to an 8.5% lower probability of replying in the top category.

## Calculating probabilities

What is the probability of the following individual reporting the highest effort category:

A male individual in good health on a permanent contract who isn’t a trade union member working in a firm of 13 has a wage is equal to the firm average:
*Relwfirm=0, Male=1, Health=1, Perm=1, Tumem=0, Emp=13, Empsq=169*

$$
\begin{aligned}
prob\left(y=0|\boldsymbol{x}\right) &= \Phi\left(-\boldsymbol{x}'\boldsymbol{\beta}\right) \\
prob\left(y=1|\boldsymbol{x}\right) &= \Phi\left(\mu_1-\boldsymbol{x}'\boldsymbol{\beta}\right)-\Phi\left(-\boldsymbol{x}'\boldsymbol{\beta}\right) \\
prob\left(y=2|\boldsymbol{x}\right) &= \Phi\left(\mu_2-\boldsymbol{x}'\boldsymbol{\beta}\right)-\Phi\left(\mu_1-\boldsymbol{x}'\boldsymbol{\beta}\right) \\
prob\left(y=3|\boldsymbol{x}\right) &= \Phi\left(\mu_3-\boldsymbol{x}'\boldsymbol{\beta}\right)-\Phi\left(\mu_2-\boldsymbol{x}'\boldsymbol{\beta}\right) \\
prob\left(y=4|\boldsymbol{x}\right) &= 1-\Phi\left(\mu_3-\boldsymbol{x}'\boldsymbol{\beta}\right)
\end{aligned}
$$

**NEED TO USE COEFFICIENTS**

$$prob\left(y=4|x\right)=1-\Phi\left[\hat{\mu}_4-\left(\begin{array}{l}\hat{\beta}_1 Relwfirm+\hat{\beta}_2 Male+\hat{\beta}_3 White+\hat{\beta}_4 Health+\hat{\beta}_5 Perm\\ +\hat{\beta}_6 Tumem+\hat{\beta}_7 Emp+\hat{\beta}_8 Empsq\end{array}\right)\right]$$

$$prob\left(y=4|x\right)=1-\Phi\left[0.388-\left(\begin{array}{l}0.276(0)+-0.255(1)+-0.197(1)+0.018(1)+0.097(1)\\ +0.136(0)+-0.00008(13)+0.000000006(169)\end{array}\right)\right]$$

$$prob\left(y=4|x\right)=1-\Phi\left[0.388--0.338\right]$$

from above z=0.726, so $\phi(z)=0.76608$

$$prob\left(y=4|x\right)=1-0.76608=0.2339$$

```
browse p4 if relwfirm==0 & male==1 & white==1 & health==1 & perm==1 &
tumem==0 & emp==13 & empsq==169
```

# NOTE

- STATA takes longer to calculate marginal effects (when *nose* is not applied) than other packages such as LIMDEP;

–This is more problematic from a research perspective. For instance in the above example WERS has info on employers and employees so it is possible to model intra-firm effects using a random effects ordered probit – takes hours to converge.

# Censored Dependent Variables

Focus on **continuous** variables and how to model economic relationships when censoring occurs in the data.

In particular the focus will be upon:

- **Truncation**

  This occurs when trying to infer the characteristics of a population from a sample which is drawn from a restricted part of the underlying population – don't observe the $y$ or **X**'s.

  Should be contrasted with **CENSORING**

- **Censoring**

  This is common in micro datasets and occurs when the dependent variable $y$ **ONLY** is censored NOT the **X**'s.

Examples (amongst many others) of censored dependent variables which have appeared in the literature are:

1. Household purchases of durable goods [Tobin (1958), *Econometrica*];

2. The number of hours worked by women in the labour market [Quester and Greene (1982), *Social Science Quarterly*];

3. Debt accumulation and financial expectations [Brown, Garino, Taylor and Wheatley Price (2005), *Economic Inquiry*].

Each of these studies analyses a dependent variable which is truncated for a significant fraction of the sample.

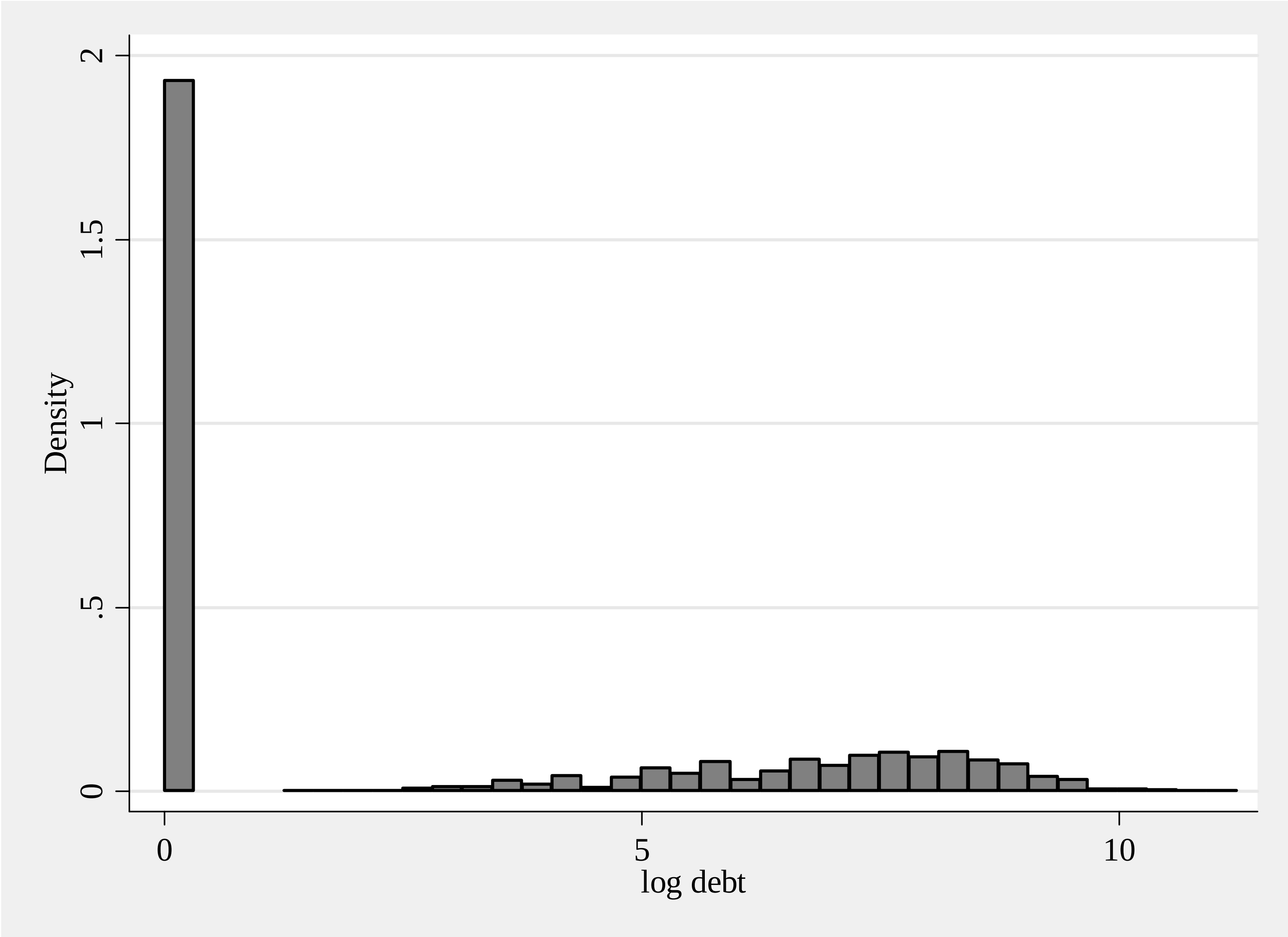
## Example of the Tobit Model – Modelling debt

Model debt using UK data from the 2000 British Household Panel Survey (*BHSP*), which consists of 3,579 individuals so $i=1,2,\ldots,3,579$.

The *BHPS* is a random sample survey, carried out by the *Institute for Social and Economic Research*, of each adult member from a nationally representative sample. For Wave one, interviews were conducted during the autumn of 1991. The same individuals are re-interviewed in successive waves – the latest available being wave twelve, collected in 2002.

In 2000, respondents were asked: *how much in total do you owe*?

| VARIABLE | DESCRIPTION |
|---|---|
| ln*Debt* | Log total amount of debt reported by the individual |
| *Age* | Age of the individual at date of interview |
| ln*Saving* | Log amount saved each month |
| ln*Income* | Log usual gross monthly pay in current job |
| ln*Wealth* | Log (investments+housevalue+windfalls+unearned income) |
| *Marrried* | Dummy variable (0/1) equals 1 if married or cohabiting |
| *Employed* | Dummy variable (0/1) equals 1 if employed |
| *Degree* | Dummy variable (0/1) equals 1 if highest qualification is a degree |
| *A'level* | Dummy variable (0/1) equals 1 if highest qualification is A'level |
| *O'level* | Dummy variable (0/1) equals 1 if highest qualification is O'level |
| *Male* | Dummy variable (0/1) equals 1 if individual is male |
| *FEI* | Financial expectations index 0=pessimistic; 1=no change; 2=optimistic |


2
1.5
1
.5
0
Density
0
5
10
log debt

## EXAMPLE *.do Tobit regression.

```
#delimit;
clear;
set mem 400m;
set mat 800;
set more off;
use "E:\karl’s files\stata\L3-4.dta";


/***Tobit model***/
tobit ldebt age lsav linc lhwealth marr emp degree alevel olevel male
ind, ll(0);
mfx compute;
predict pldebt;
gen pdebty=exp(pldebt);
```

## RESULTS FILE

```
tobit ldebt age lsav linc lhwealth marr emp degree alevel olevel male ind, ll(0);

Tobit estimates                                   Number of obs   =       3579
                                                  LR chi2(11)     =     407.08
                                                  Prob > chi2     =     0.0000
Log likelihood = -6071.7422                       Pseudo R2       =     0.0324
------------------------------------------------------------------------------
       ldebt |      Coef.   Std. Err.       t     P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.1166167   .0149277     -7.81   0.000    -.1458843   -.0873491
        lsav |  -.1110862   .0649746     -1.71   0.087    -.2384774    .0163049
        linc |   .4988664   .1199528      4.16   0.000     .2636834    .7340494
    lhwealth |  -.3126518   .0345654     -9.05   0.000    -.3804217   -.2448819
        marr |   -.484894   .3150764     -1.54   0.124    -1.102642     .132854
         emp |   1.087713   .3329233      3.27   0.001     .4349735    1.740452
      degree |   .1925321   .3877163      0.50   0.620    -.5676358       .9527
      alevel |   .4574903   .2892543      1.58   0.114    -.1096301    1.024611
      olevel |  -.0971489   .3509876     -0.28   0.782    -.7853053    .5910076
        male |   .4997765   .3007312      1.66   0.097    -.0898458    1.089399
         ind |   1.187578   .2537143      4.68   0.000     .6901385    1.685018
       _cons |  -.2806632   1.050006     -0.27   0.789    -2.339336     1.77801
-------------+----------------------------------------------------------------
         _se |   7.121927   .1566116           (Ancillary parameter)
------------------------------------------------------------------------------
  Obs. summary:       2158  left-censored observations at ldebt<=0
                      1421     uncensored observations
```

```
mfx compute;

Marginal effects after tobit
      y  = Fitted values (predict)
         = -1.2557238
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
     age |  -.1166167      .01493   -7.81   0.000  -.145874 -.087359   44.7913
    lsav |  -.1110862      .06497   -1.71   0.087  -.238434  .016262   1.80164
    linc |   .4988664      .11995    4.16   0.000   .263763   .73397   6.16907
lhwealth |  -.3126518      .03457   -9.05   0.000  -.380399 -.244905   2.64378
   marr*|   -.484894      .31508   -1.54   0.124  -1.10243  .132644   .694049
    emp*|   1.087713      .33292    3.27   0.001   .435195  1.74023    .65074
 degree*|   .1925321      .38772    0.50   0.619  -.567378  .952442   .175189
 alevel*|   .4574903      .28925    1.58   0.114  -.109438  1.02442   .406259
 olevel*|  -.0971489      .35099   -0.28   0.782  -.785072  .590774   .214864
   male*|   .4997765      .30073    1.66   0.097  -.089646   1.0892   .395641
     ind |   1.187578      .25371    4.68   0.000   .690307  1.68485   1.20397
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

## What do the marginal effects (Coefficients) mean from the Tobit?

i. A 1% increase in savings reduces debt by 11.1%

ii. If income goes up by 1% then debt increases by 49.9%

iii. Individuals with a degree have 19.3% more debt than those with no qualifications

*How much debt does the following individual have?*

a male individual aged 34 – *Male=1*;
with no savings or wealth – ln*Saving=0*; ln*Wealth=0;*
income of £736.54 – ln*inc*=6.602
employed – *Emp=1;*
married – *Marr=1*;
with no education – *Degree=0*, *O'level=0*, *A'level=0*;
who is financially optimistic – *FEI=2*?

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 Age_i + \hat{\beta}_2 \ln Saving_i + \hat{\beta}_3 \ln Income_i + \hat{\beta}_4 \ln Wealth_i + \hat{\beta}_5 Married_i$$

$$+ \hat{\beta}_6 Employed_i + \hat{\beta}_7 Degree_i + \hat{\beta}_8 A' level_i + \hat{\beta}_9 O' level_i + \hat{\beta}_{10} Male_i + \hat{\beta}_{11} FEI_i$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1(34) + \hat{\beta}_2(0) + \hat{\beta}_3(6.602) + \hat{\beta}_4(0) + \hat{\beta}_5(1) + \hat{\beta}_6(1) + \hat{\beta}_7(0) + \hat{\beta}_8(0) + \hat{\beta}_9(0) + \hat{\beta}_{10}(1) + \hat{\beta}_{11}(2)$$

$$\hat{y}_i = -0.2807 + -0.1166(34) + 0 + 0.4989(6.602) + 0 + -0.4849(1) + 1.0877(1) + 0 + 0$$

$$+ 0 + 0.49981(1) + 1.1876(2)$$

$$\hat{y}_i = 2.5256 = £12.49$$

```
browse pdebty pldebt debty ldebt if age==34 & lsav==0 & linc>6.6 &
linc<6.602 & lhwealth==0 & marr==1 & emp==1 & alevel==0 & degree==0 &
olevel==0 & male==1 & ind==2
```

*How we calculate the probability that an individual has between 0 and £1,000 debt?*

$$prob(l_i, u_i) = prob\left(l_i < \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i < u_i\right) = \Phi\left(\frac{u_i - \hat{y}_i}{\sigma}\right) - \Phi\left(\frac{l_i - \hat{y}_i}{\sigma}\right)$$

In the above example:

$$prob(0, 6.9077) = \Phi\left(\frac{6.9077 - \hat{y}_i}{7.1219}\right) - \Phi\left(\frac{0 - \hat{y}_i}{7.1219}\right)$$

```
/***Probability that an individual has between 0 and £1000***/
predict p, pr(0,6.9077);
gen ste=7.121927;
gen lower=norm((0-pldebt)/ste);
gen upper=norm((6.9077-pldebt)/ste);
gen prob=upper-lower;
sum p prob;
```

*What about the probability that the same individual (as defined above) has debt between £1,000 and £5,000?*

We know from above $\hat{y}_i = 2.5256$

$$prob(0, 6.9077) = \Phi\left(\frac{8.5172 - 2.5256}{7.1219}\right) - \Phi\left(\frac{6.9077 - 2.5256}{7.1219}\right) = 0.0691$$

$\therefore$ A 7% probability