Lecture Notes on Advanced Econometrics

Lecture 13: Dummy and Qualitative Dependent Variables

In this lecture, we study econometric methods when the dependent variable is a limited dependent variable. The dependent variable can be limited variables in various ways. In this lecture, we study when the dependent variable is observed as a binary variable and when the dependent variable is categorical.

Binary Response Models: Linear Probability Model, Logit, and Probit

Consider a general $k + 1$ variable equation:

$$ y_i^* = \beta' x_i + u_i \quad \text{for } i = 1, \ldots, n. $$

The problem is that we do not observe $y_i^*$. Instead, we observe the binary variable:

$$ y_i = \begin{cases} 
1 & \text{if } y_i^* = \beta' x_i + u_i > c \\
0 & \text{otherwise}
\end{cases} $$

Here, the probability of $y_i$ being 1 is equal to the probability of $y_i^*$ being larger than a constant $c$, where $c$ is a threshold. The threshold is easily converted into 0 by adjusting the constant term in the model. Thus, for simplicity, we rewrite the condition as

$$ y_i = \begin{cases} 
1 & \text{if } y_i^* = \beta' x_i + u_i > 0 \\
0 & \text{otherwise}
\end{cases} . $$

Note that the probability of the observed $y_i$ being one can be written by using $y_i^*$ as

$$ \Pr(y_i = 1) = \Pr(y_i^* > 0) = \Pr(u_i > -\beta' x_i) . $$
We assume that the error term \( u_i \) has a cumulative distribution function of \( F(u_i) \) and where \( f(u_i) \) is the probability density function of \( F(u_i) \). Then, we have

\[
\Pr(u_i > -\beta' x_i) = 1 - F(-\beta' x_i) = F(\beta x_i),
\]

because the probability of \( u_i \) being larger than \(-\beta' x_i\) is the shaded area the left panel of Figure 1. If the distribution has a symmetric shape, the shaded areas in the left panel are equal to the shaded area in the right panel of Figure 1.

Figure 1.

If we replace the cumulative distribution function with the standard normal distribution function, then we have the Probit model.

**Probit:** replace \( F(\beta x_i) \) with \( \Phi(\beta x_i) \)

\( \Phi(\beta x_i) \) is the standard normal distribution.

\[
\Phi(\beta' x) = \int_{-\infty}^{\infty} \phi(z) dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz
\]

If we replace the cumulative distribution function with the logistic distribution, then we have the Logit model.
Logit: replace $F(\beta x_i)$ with $\Lambda(\beta x_i)$

$\Lambda(\beta x_i)$ is the logistic cumulative distribution function.

$$\Lambda(\beta'x) = \frac{e^{\beta'x}}{1 + e^{\beta'x}}$$

In the observations, we have $n$ cases of zeros and ones with probability of $F(\beta x_i)$ for ones and $1-F(\beta x_i)$ for zeros. Thus the joint probability or the likelihood function is

$$\Pr(y_1 = 0, y_2 = 1, \ldots, y_n = 0) = \prod_{i=1}^{n} [1 - F(\beta' x_i)] \prod_{i=1}^{n} F(\beta' x_i)$$

or

$$L = \prod_{i=1}^{n} [1 - F(\beta' x_i)]^{y_i} \cdot [F(\beta' x_i)]^{1-y_i}.$$

Taking logs, we have

$$\ln L = \sum_{i} \left[ (1-y_i) \ln(1-F(\beta' x_i)) + y_i \ln F(\beta' x_i) \right]$$

This is the log likelihood. By differentiating $\ln L$ with respect to $\beta$, we get

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i} \left\{ - \frac{f(\beta' x_i) x_i (1-y_i)}{1-F(\beta' x_i)} + \frac{f(\beta' x_i) x_i y_i}{F(\beta' x_i)} \right\}$$
\[
= \sum_i \left\{ \frac{[y_i - F(\beta' x_i)]}{[1-F(\beta' x_i)]F(\beta' x_i)} f(\beta' x_i) x_i \right\}
\]

The maximum likelihood estimators are obtained at \( \frac{\partial \ln L}{\partial \beta} = 0 \).

Thus, we estimate this likelihood function to get \( \beta \) that maximizes the likelihood function.

**Probit:** \( \ln L = \sum_{y_i=0} \ln(1-\Phi(\beta' x_i)) + \sum_{y_i=1} \ln \Phi(\beta' x_i) \)

**Logit:** \( \ln L = \sum_{y_i=0} \ln(1-\Lambda(\beta' x_i)) + \sum_{y_i=1} \ln \Lambda(\beta' x_i) \)

**Simple Example of Logit model**

Consider the following model:

\[
y_i^* = \alpha + \beta d_i + u_i
\]

\( y = 1 \) if \( y_i^* > 0 \)

\( y = 0 \) otherwise

As we studied in the class, the log likelihood function is

\[
\ln L = \sum_i [(1-y_i) \ln(1-F(\beta' x_i)) + y_i \ln F(\beta' x_i)]
\]

(1)

Now, replace the cumulative function with the logistic function:

\[
F(\beta' x) = \Lambda(\beta' x) = \frac{e^{\beta' x}}{1 + e^{\beta' x}} = \frac{e^\alpha + \beta d}{1 + e^{\alpha + \beta d}}
\]

(a) Write down the log likelihood function (1) with the logistic function.

(b) Suppose you have the data of 10 observations:

<table>
<thead>
<tr>
<th>( d )</th>
<th>0</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Show that in this case, the log likelihood function becomes
\[ \ln L = 7\alpha - 10\ln(1 + e^\alpha) \] (2)

(c) By differentiating the log likelihood function (2) w.r.t. \( \forall \), find \( \forall \) that maximizes the log likelihood function. (Hint: \( \frac{\partial e^\alpha}{\partial \alpha} = e^\alpha \))

(d) By using the data called MLE.dta, estimate the model and confirm your answer.

(e) Suppose you have the data of 10 observations:

<table>
<thead>
<tr>
<th>( d )</th>
<th>( Y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

Find \( \hat{\alpha} \) and \( \hat{\beta} \) by the same method in (a)–(d). Show me how you obtain the estimators.

The Likelihood Ratio Test

\[ LR = -2[\ln L_r - \ln L_{ur}] \]

This is like a F-test for the maximum likelihood estimation. Note that in maximum likelihood estimation, we want to maximize the likelihood, instead of minimizing (the least square). Thus the larger the better (remember however that the likelihood is a negative number), thus the likelihood of the unrestricted model, \( \ln L_{ur} \), will be larger than the restricted model, \( \ln L_r \). Thus the inside the bracket will be negative, and \( LR \) will be positive. If \( LR \) is a large number: if there is a significant difference in \( \ln L_r \) and \( \ln L_{ur} \), then the additional variables in the unrestricted model are jointly significant.
Measuring the Goodness of Fit

The percent correctly predicted: define

\[ \hat{y}_i = 1 \text{ if } y_i^* > 0.5 \quad \text{and} \quad \hat{y}_i = 1 \text{ otherwise} \]

and create a 2 by 2 matrix. And calculate a percentage of correct prediction on both \( y_i = 1 \) and \( y_i = 0 \).

Alternatively you can get a pseudo R-squared:

\[
\text{Pseudo R-squared} = 1 - \frac{\ln L}{\ln L_0}
\]

\( \ln L_0 \) is the log likelihood from a model with the constant term only.

Example 13-1: Food Aid Targeting in Ethiopia (Jayne et al., 2002)

```
. ***** WEREDA LEVEL ANALYSIS *****;
. *** Food for Free ****;

. probit wvfreed;

Probit estimates                          Number of obs = 343
                                   LR chi2(0) = -0.00
                                   Prob > chi2 = .

Log likelihood = -200.44456           Pseudo R2 = -0.0000

------------------------------------------------------------------------------
   wvfreed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     _cons |  -.6093778   .0724429    -8.41   0.000    -.7513633   -.4673923
------------------------------------------------------------------------------
```


. probit wvfreed teg am oro other lpcninc elev
> shrtplot overplot diseplot;

Probit estimates

Number of obs = 343
LR chi2(9) = 79.36
Prob > chi2 = 0.0000
Log likelihood = -160.76343 Pseudo R2 = 0.1980

------------------------------------------------------------------------------
|                | Coef.   | Std. Err.  | z    | P>|z|  | [95% Conf. Interval] |
|----------------|---------|------------|------|------|---------------------|
| Tigray         | 1.467662| 0.3326602  | 4.41 | 0.000| 0.8156605            |
| Amhara         | 0.1965025| 0.2465459  | 0.80 | 0.425| -0.2867187           |
| Oromiya        | -0.1079401| 0.2379339  | -0.45| 0.650| -0.5742821           |
| Other          | 0.8938703| 0.3946693  | 2.26 | 0.024| 0.1203327            |
| Ln(pc inc)     | -0.4626233| 0.1515389  | -3.05| 0.002| -0.7596341           |
| Elevation      | 0.0340291| 0.017033   | 2.00 | 0.046| 0.000645             |
| Shortage       | 0.0200977| 0.0050222  | 4.00 | 0.000| 0.0102542            |
| Flood          | 0.012639 | 0.0071791  | 1.76 | 0.078| -0.0014317           |
| Insect         | 0.0070072| 0.0078859  | 0.89 | 0.374| -0.0084488           |
| _cons          | 0.8524934| 0.8673863  | 0.98 | 0.326| -0.8475525           |
------------------------------------------------------------------------------

** This is a Wald test **************;

. test shrtplot overplot diseplot;

( 1) shrtplot = 0.0
( 2) overplot = 0.0
( 3) diseplot = 0.0

ch2( 3) = 17.59
Prob > ch2 = 0.0005

** Log Likelihood Test, obtaining the log likelihood from unrestricted model;

. lrtest, saving(0);
** Obtaining the percent correctly predicted *****

. predict y1;
(option p assumed; Pr(wvfreed))
. gen y1_one=(y1>0.5);
. tab wvfreed y1_one, row;

<table>
<thead>
<tr>
<th>y1_one</th>
</tr>
</thead>
<tbody>
<tr>
<td>wvfreed</td>
</tr>
<tr>
<td>-----------+----------------------+----------</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*** Restricted Model;

. probit wvfreed teg am oro other lpcninc elev;

Probit estimates
Number of obs = 343
LR chi2(6) = 60.99
Prob > chi2 = 0.0000
Log likelihood = -169.9489
Pseudo R2 = 0.1521
------------------------------------------------------------------------------
| wvfreed | Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|---------|-------|-----------|-------|-----|-----------------------|
| teg     | 1.679131 | .3217638 | 5.22  | 0.000 | 1.048485 - 2.309776   |
| am      | .349624  | .2339933  | 1.49  | 0.135 | -.1089944 - .8082424  |
| oro     | -.0745053 | .2318984 | -0.32 | 0.748 | -.5290178 .3800072    |
| other   | .8442578  | .3812492  | 2.21  | 0.027 | .0970232 1.591492     |
| lpcninc | -.5494286 | .146091 | -3.76 | 0.000 | -.8357616 -.2630956   |
| elev    | .0264549  | .0163385  | 1.62  | 0.105 | -.0055679 .0584778    |
Marginal Effects

The coefficients of Probit or Logit in likelihood function do not represent changes in probabilities. To get effects on marginal probability, we need to transform the estimated coefficient.

The probability model is

\[ E[y] = 0 \ [1 - F(\beta' x_i)] + 1 \ [F(\beta' x_i)] \]
\[ = F(\beta' x_i) \]

Thus, the marginal changes in the probability is, by using the chain rule,

\[ \frac{\partial E[y]}{\partial x_k} = \left( \frac{d F(\beta' x)}{d(\beta' x_k)} \right) \frac{d(\beta' x_k)}{dx_k} \]
\[ = f(\beta' x_i) \beta_k . \]

For Probit and Logit, we have

Probit:\n\[ \frac{\partial E[y]}{\partial x_k} = \phi(\beta' x) \beta_k \]

Logit:\n\[ \frac{\partial E[y]}{\partial x_k} = \Lambda (\beta' x)(1 - \Lambda(\beta' x)) \beta_k \]

Evaluate at means:
Probit: \[
\frac{\partial E[y]}{\partial x_k} = \phi(\hat{\beta}' \bar{x}) \hat{\beta}_k
\]

Logit: \[
\frac{\partial E[y]}{\partial x_k} = \Lambda(\hat{\beta}' \bar{x}) \left(1 - \Lambda(\hat{\beta}' \bar{x})\right) \hat{\beta}_k
\]

Changes in Probability when a Change in \( x \) is Not So Marginal

Because Probit and Logit are no-linear model, a marginal change (which is a linear approximation at some point) can be misleading. Thus, we need to conduct a simulation. Suppose we want to know a change in the probability of \( y_i = 1 \) when \( x_s \) changes from \( a \) to \( b \):

\[
\Pr (y_i = 1 | x_s = b) - \Pr (y_i = 1 | x_s = a)
\]

See the next example.
Example 13-2: Food Aid Targeting in Ethiopia (Jayne et al., 2002)

dprobit: Marginal Probability Model

.dprobit wvfreed teg am oro other lpcninc elev
>       shrtplot overplot diseplot;

Iteration 0:  log likelihood = -200.44456
Iteration 1:  log likelihood = -161.75807
Iteration 2:  log likelihood = -160.76658
Iteration 3:  log likelihood = -160.76343

Probit estimates

Number of obs = 343
LR chi2(9) = 79.36
Prob > chi2 = 0.0000
Log likelihood = -160.76343
Pseudo R2 = 0.1980

------------------------------------------------------------------------------
 wvfreed |      dF/dx   Std. Err.      z    P>|z|     x-bar  [ 95% C.I. ]
---------+--------------------------------------------------------------------
    teg* |   .5347784   .1071274     4.41   0.000   .090379   .324813  .744744
      am* |   .0632314    .081563     0.80   0.425   .268222  -.096629  .223092
     oro* |  -.0334596   .0731175    -0.45   0.650   .405248  -.176767  .109848
     other* |   .331359   .1537268     2.26   0.024   .046647    .03006  .632658
    lpcninc |  -.1444777   .0471902    -3.05   0.002   5.80889  -.236969 -.051987
     elev |   .0106273   .0053209     2.00   0.046   20.4433   .000199  .021056
  shrtplot |   .0062765   .0015886     4.00   0.000    7.34191  .003163  .00939
  overplot |   .0039472   .0022367     1.76   0.078   4.51159  -.000437  .008331
 diseplot |   .0021883   .0024627     0.89   0.374    7.34973  -.002638  .007015
------------------------------------------------------------------------------

                   obs. P |   .271137
                  pred. P |   .2420307  (at x-bar)

(*) dF/dx is for discrete change of dummy variable from 0 to 1
    z and P>|z| are the test of the underlying coefficient being 0
**** Predict at means;
. predict ymean;
(option p assumed; Pr(wvfreed))

*** Get summary of log(per capita income);
. su lpcninc, d;

(mean) lpcninc

-------------------------------------------------------------
Percentiles      Smallest
1%     4.339287       3.519491
5%     4.836071       3.980483
10%     5.108971       4.324318       Obs                 343
25%     5.446068       4.339287       Sum of Wgt.         343
50%     5.863084                      Mean           5.808894
     Largest                   Std. Dev.      .5785665
75%     6.182746       6.987036
90%     6.546782       7.029579       Variance       .3347392
95%     6.76193       7.078065       Skewness      -.3397844
99%     6.987036       7.130215       Kurtosis      3.425768
*** A Simulation in Pr(y) when lnpcninc changes from 25th to 75th percentile;

** Set the lnpcninc at 25th percentile;
. replace lpcninc=5.446;
(343 real changes made)

** Predict at Pr(y) at 25th, holding other variables at means;
. predict yat25;
(option p assumed; Pr(wvfreed))

** Set the lnpcninc at 75th percentile;
. replace lpcninc=6.813;
(343 real changes made)

** Predict at Pr(y) at 75th, holding other variables at means;
. predict yat75;
(option p assumed; Pr(wvfreed))

. su ymean yat25 yat75;

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>ymean</td>
<td>343</td>
<td>.2713801</td>
<td>.2179637</td>
<td>.0246233</td>
<td>.9457644</td>
</tr>
<tr>
<td>yat25</td>
<td>343</td>
<td>.312233</td>
<td>.2115653</td>
<td>.0845005</td>
<td>.9469762</td>
</tr>
<tr>
<td>yat75</td>
<td>343</td>
<td>.1599189</td>
<td>.1798231</td>
<td>.0223308</td>
<td>.8373954</td>
</tr>
</tbody>
</table>

End of Example 12
Fixed Effects Logit Model

A fixed effects Logit model that accounts for the heterogeneity is

\[
\Pr(y_{it} = 1) = \frac{e^{\alpha_i + \beta x_{it}}}{1 + e^{\alpha_i + \beta x_{it}}},
\]

where \( \alpha_i \) is the fixed effect for observation \( i \). For simplicity, we only consider cases for \( T = 2 \). The unconditional likelihood is

\[
L = \prod_i \Pr(Y_{i1} = y_{i1}) \Pr(Y_{i2} = y_{i2}),
\]

where \( y_{it} \) is the value that \( Y_{it} \) takes at time \( t \) for observation \( i \). (The likelihood is the product of the probabilities because the observations are independent.) The sum of \( Y_{i1} \) and \( Y_{i2} \) can take three values: 0, 1, and 2. When the sum is zero, there is only one combination \( (Y_{i1}, Y_{i2}) = (0, 0) \). When the sum is two, there is also one combination: \( (Y_{i1}, Y_{i2}) = (1, 1) \). In these two cases, we do learn nothing about the attributes for \( x_i \) on \( Y_{it} \) because there is no variation in \( Y_{it} \). Observations with these combinations will be dropped from the analysis.

When the sum is one, there are two possibilities: \( (Y_{i1}, Y_{i2}) = (1, 0) \) and \( (0, 1) \). Thus, we have

\[
\Pr(0,1 \mid sum = 1) = \frac{\Pr(0,1)}{\Pr(0,1) + \Pr(1,0)}, \text{ or}
\]

\[
\Pr(1,0 \mid sum = 1) = \frac{\Pr(1,0)}{\Pr(0,1) + \Pr(1,0)}.
\]

The conditional probability by using logit functions is

\[
\frac{\Pr(1) \Pr(0)}{\Pr(0) \Pr(1) + \Pr(1) \Pr(0)}
\]
\[
\frac{e^{\alpha_i + \beta x_{i1}}}{1 + e^{\alpha_i + \beta x_{i1}}} \cdot \frac{1}{1 + e^{\alpha_i + \beta x_{i2}}} = \frac{e^{\alpha_i + \beta x_{i1}}}{1 + e^{\alpha_i + \beta x_{i1}}} + \frac{e^{\alpha_i + \beta x_{i1}}}{1 + e^{\alpha_i + \beta x_{i1}}} \cdot \frac{1}{1 + e^{\alpha_i + \beta x_{i2}}}
\]

\[
= e^{\alpha_i + \beta x_{i1}} + e^{\alpha_i + \beta x_{i1}} = e^{\beta x_{i1}} + e^{\beta x_{i1}} = \frac{1}{e^{\beta x_{i2} - x_{i1}}} + 1
\]

\[
= 1 - \Lambda[\beta'(x_{i2} - x_{i1})]
\]

Thus, in the fixed effects logit model, the fixed effect, \(\alpha_i\), is eliminated. Similarly, the

\[
\frac{Pr(0) Pr(1)}{Pr(0) Pr(1) + Pr(1) Pr(0)} = 1 - \frac{1}{e^{\beta'(x_{i2} - x_{i1})} + 1}
\]

\[
= \Lambda[\beta'(x_{i2} - x_{i1})]
\]

Thus, the log likelihood function for observation \(i\) is

\[
L_i = \sum [w_i \ln(\Lambda(\beta'x_i) + (1 - w_i) \ln(1 - \Lambda(x_i)))]
\]

where \(w_i = 1\) if \((Y_{i1}, Y_{i2}) = (0, 1)\) and \(w_i = 0\) if \((Y_{i1}, Y_{i2}) = (1, 0)\).

The estimated coefficients of the fixed effects logit model can be interpreted as the effects on log-odd ratio. But we are unable to find level effects of \(x_i\) on the dependent variables. This is partly the reasons that some researchers use the (liner) fixed effects model on the binary response dependent variables to control for fixed effects.
Multinomial Logit Model

The logit model for binary response can be extended for more than two outcomes that are not ordered. Let $y$ denote a random variable taking on the values $[0, 1, \ldots, J]$. Since the probabilities must sum to unity, $\Pr(y = 0|x)$ is determined once we know the probabilities for $j = 1, \ldots, J$. Therefore, the probability that $y$ takes $j$ can be written as

$$
\Pr(Y_i = j) = \frac{e^{\beta'_j x_i}}{1 + \sum_{k=1}^{J} e^{\beta'_k x_i}} \text{ for } j = 1, \ldots, J
$$

and

$$
\Pr(Y_i = 0) = \frac{1}{1 + \sum_{k=1}^{J} e^{\beta'_k x_i}} \text{ for } j = 0.
$$

Note that for a binary response case, $J = 2$, this becomes the logit model. The model implies that we can compute $J$ log-odd ratios

$$
\ln \left[ \frac{P_y}{P_{y0}} \right] = \beta'_x.
$$

This implies that the estimated coefficients of the Multinomial Logit should be interpreted as the effects on the log-off ratios of choice $j$ and the base choice, $j = 0$.

The marginal effects on the particular outcome are

$$
\frac{\partial P}{\partial x_i} = P_i [\beta_j - \sum_{k=1}^{J} P_k \beta_k].
$$

Note that this result suggests that the direction of the marginal effect could be different from the sign of the estimated coefficient (of the odd-ratio).

Quiz: Show that the sum of the marginal effects for all outcomes is zero.

(STATA 9 provides the marginal effects.)

In this example, the dependent variable is a choice variable: Not attending school, attending primary school, and attending secondary school. The focus is on orphans who lost at least one parent before reaching 15 years old. The sample in this example include girls aged 15 to 18.

```
.xi: mlogit type oneporp vdorphan onepnorp bothout age16 age17 age18  fhh
edumen eduwomen eldmale eldfmale wkmen wkwomen lsize lprop i.distcode if
sex==2, base(0)
i.distcode        _Idistcode_12-40    (naturally coded; _Idistcode_12 omitted)

Multinomial logistic regression                   Number of obs   =        384
LR chi2(88)     =     247.35
Prob > chi2     =     0.0000
Log likelihood = -279.30422                       Pseudo R2       =     0.3069

------------------------------------------------------------------------------
     type |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Primary School
    SingleOrphans|    -.67788   .7392469    -0.92   0.359    -2.126777    .7710173
   DoubleOrphans|  -1.382794   .6558163    -2.11   0.035     -2.66817   -.0974173
      One parent |   1.047395   .7657416     1.37   0.171    -.4534311    2.548221
       No parents |  -1.347218   .6773663    -1.99   0.047    -2.674832   -.0196047
    age16 |  -2.014747   .4909821    -4.10   0.000    -2.977054   -1.05244
    age17 |  -1.864725   .5329909    -3.50   0.000    -2.909368    -.8200819
    age18 |  -4.219341   .6813643    -6.19   0.000    -5.554791   -2.883892
      Female headed|   .1620776   .6449858     0.25   0.802    -1.102071    1.426226
       edumen |  -.0605027   .0656405    -0.92   0.357     -.189155    .0681504
     eduwomen |  -.1250235   .0888526    -1.41   0.159     -.2991714    .0491244
      eldmale |   1.613167   .6189864     2.61   0.009     .3999752    2.826358
     eldfmale |   .1462345   .673432    0.22   0.828     -.1173668    1.466137
------------------------------------------------------------------------------
``
| Variable          | Coefficient | Std. Error | z  | P>|z< | Lower | Upper |
|-------------------|-------------|------------|----|------|-------|-------|
| wkmen             | -0.30326    | 0.2035947  | -0.15 | 0.882 | -0.4293643 | 0.367124 |
| wkwomen           | -0.1877117  | 0.1806674  | -1.04 | 0.299 | -0.5418134 | 0.16639 |
| log(Landsize)     | 0.0680674   | 0.2684975  | 0.25  | 0.800 | -0.458178 | 0.5943128 |
| log(assets)       | 0.3132038   | 0.1841006  | 1.70  | 0.089 | -0.0476267 | 0.6740342 |
| _Idistcod-13      | 0.5352554   | 1.158086   | 0.46  | 0.644 | -1.734552 | 2.805062 |
| _Idistcod-40      | -39.98424   | 3.19e+08   | -0.00 | 1.000 | -6.26e+08 | 6.26e+08 |
| _cons             | -0.2099659  | 2.205778   | -0.10 | 0.924 | -4.533211 | 4.113279 |

Secondary School

| Variable          | Coefficient | Std. Error | z  | P>|z< | Lower | Upper |
|-------------------|-------------|------------|----|------|-------|-------|
| SingleOrphans     | -0.96072    | 0.6962642  | -1.38 | 0.168 | -2.325325 | 0.403908 |
| Doubleorphans     | -2.159722   | 0.6353289  | -3.40 | 0.001 | -3.404944 | -0.914506 |
| One parent        | -0.9110557  | 0.816869   | -1.12 | 0.265 | -2.512089 | 0.689978 |
| No parents        | -1.975631   | 0.6386655  | -3.09 | 0.002 | -3.227393 | -0.72387 |
| age16             | -0.4085832  | 0.5078172  | -0.80 | 0.421 | -1.403886 | 0.596322 |
| age17             | -0.093862   | 0.5384703  | -0.17 | 0.862 | -1.149244 | 0.9615205 |
| age18             | -0.7141053  | 0.5384291  | -1.33 | 0.185 | -1.769407 | 0.3411964 |
| Female headed     | 0.5078927   | 0.5961982  | 0.85  | 0.394 | -0.6606342 | 1.67642 |
| edumen            | -0.0067266  | 0.0619717  | -0.11 | 0.914 | -0.128189 | 0.1147357 |
| eduwomen          | 0.115019    | 0.0816583  | 1.41  | 0.159 | -0.0450284 | 0.2750663 |
| eldmale           | 0.9500993   | 0.609837   | 1.56  | 0.119 | -0.2451593 | 2.145358 |
| elfmale           | 0.6985087   | 0.6432948  | 1.09  | 0.278 | -0.5623261 | 1.959343 |
| wkmen             | -0.1848425  | 0.1923146  | -0.96 | 0.336 | -0.5617723 | 0.1920873 |
| wkwomen           | 0.5618054   | 0.1758088  | -3.20 | 0.001 | -0.9063843 | -0.2172265 |
| log(Landsize)     | 0.4230313   | 0.2572931  | 1.64  | 0.100 | -0.081254 | 0.9273165 |
| log(assets)       | 0.4838305   | 0.1805556  | 2.68  | 0.007 | -0.129948 | 0.837713 |
| _Idistcod-13      | -0.2492886  | 1.210194   | -0.21 | 0.837 | -2.621225 | 2.122648 |
| _Idistcod-40      | 0.3812108   | 1.052142   | -0.36 | 0.717 | -2.443371 | 1.680949 |
| _cons             | -5.631196   | 2.233789   | -2.52 | 0.012 | -10.00934 | -1.253051 |

(Outcome type==0 is the comparison group)
Lecture 14: Censored and Truncated Dependent Variables

Example 1: A Simple Example of Truncated Mean

\( y \) has a uniform distribution:
\[ f(y) = 1, \quad 0 \leq y \leq 1 \]

For truncation at \( y = 1/3 \)
\[
\begin{align*}
&\quad f(y \mid y > 1/3) = \frac{f(y)}{Pr(y > 1/3)} = \frac{1}{2/3} = \frac{3}{2} \\
\end{align*}

Thus
\begin{align*}
E[y \mid y > 1/3] &= \int_{1/3}^{1} y f(y \mid y > 1/3) \, dx = \int_{1/3}^{1} y (3/2) \, dx = (3/2) \int_{1/3}^{1} y \, dx \\
&= (3/2) [(1/2)(1-1/3)] = 2/3
\end{align*}

End of Example 1

If the truncation is from below, the mean of the truncated variable is greater than the mean of the original one. And if the truncation from above, the mean of the truncated variable is smaller than the original one. Thus, \( E[y \mid y < a] < \mu \) and \( E[y \mid y > a] > \mu \).

Censored Dependent Variables

Suppose that a continuous random variable, \( y \), is censored (or has a corner solution) at \( a \) and is observed above \( a \):
\[ y = \begin{cases} 
    y^* & \text{if } y^* > a \\
    a & \text{otherwise}
\end{cases} \quad \text{where } y^* = X'\beta + u \]

and
\[ u \sim N[0, \sigma^2]. \]

Note here that we can rewrite \( y^* > a \) as
\[
\begin{align*}
X'\beta + u &> a \\ u &> a - X'\beta
\end{align*}
\]

Thus, the probability of \( y \) being larger than \( a \) can be write as:
\[ \Pr(y > a) = \Pr\left(\frac{y}{\sigma} > \frac{a}{\sigma}\right). \]

\[ = \Pr\left(\frac{u}{\sigma} > \frac{a - X'\beta}{\sigma}\right) \]

\[ = 1 - \Phi((a - X'\beta)/\sigma) \]

\[ = \Phi((X'\beta - a)/\sigma). \]

The probability of \( y \) being smaller than \( a \) is just \( 1 - \Phi((X'\beta - a)/\sigma) = \Phi((a - X'\beta)/\sigma) \).

Also note that the conditional distribution of \( u \) can be written as

\[ f(u \mid u > a - X'\beta) = \sigma f\left(\frac{u}{\sigma} > \frac{a - X'\beta}{\sigma}\right) \]

\[ = \sigma \frac{\phi(u/\sigma)}{\Pr\left(\frac{u}{\sigma} > \frac{a - X'\beta}{\sigma}\right)} \]

\[ = \frac{\phi(u/\sigma)}{\Phi((X'\beta - a)/\sigma)}. \]

Now, let’s consider the expected value of \( y \) where \( y \) is larger than \( a \):

\[ E(y \mid y > a) = X'\beta + E[u \mid u > a - X'\beta], \]

where

\[ E(u \mid u > a - X'\beta) = \sigma \int_{(a - X'\beta)/\sigma}^{\infty} \frac{(u/\sigma) f(u/\sigma)}{\Pr\left(\frac{u}{\sigma} > \frac{a - X'\beta}{\sigma}\right)} d(u/\sigma) \]

\[ = \frac{\sigma}{\Phi((X'\beta - a)/\sigma)} \int_{(a - X'\beta)/\sigma}^{\infty} (u/\sigma) \phi(u/\sigma) d(u/\sigma). \quad (1) \]

Note here that

\[ \frac{\partial \phi(z)}{\partial z} = \frac{\partial}{\partial u} \left( \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \right) \]

\[ = \frac{1}{\sqrt{2\pi}} (-z) \exp\left(-\frac{z^2}{2}\right) \]
Thus,
\[ \int u \phi(z)\,dz = -\phi(z). \]

Therefore, (1) can be written as
\[
E(u \mid u > a - X'\beta) = \frac{\sigma}{\Phi((X'\beta - a) / \sigma)} [-\phi(u / \sigma)]_{(a-X'\beta)/\sigma} \\
= \frac{\sigma}{\Phi((X'\beta - a) / \sigma)} [-\phi(\infty) + \phi((a - X'\beta) / \sigma)] \\
= \frac{\sigma \phi((a - X'\beta) / \sigma)}{\Phi((X'\beta - a) / \sigma)}.
\]

Because the normal density function is symmetric, this can be written as;
\[
E(u \mid u > a - X'\beta) = \frac{\sigma \phi((X'\beta - a) / \sigma)}{\Phi((X'\beta - a) / \sigma)} = \sigma \lambda((X'\beta - a) / \sigma).
\]

The \( \lambda(\cdot) \) is called the inverse Mills ratio. Notice that if the cut-off point is at zero, \( a = 0 \), then we have:
\[
E(y \mid y > 0) = X'\beta + E[u \mid u > -X'\beta] \\
= X'\beta + \frac{\sigma \phi(X'\beta / \sigma)}{\Phi(X'\beta / \sigma)} \\
= X'\beta + \sigma \lambda(X'\beta / \sigma).
\]

**Marginal effects of \( x_i \) on \( y_i \),**

The conditional expected value of \( y \) on \( X \) is
\[
E(y \mid X) = \Pr(y = 0 \mid X) \times 0 + \Pr(y > 0 \mid X) \times E(y \mid X, y > 0) \\
= \Pr(y > 0 \mid X) \times E(y \mid X, y > 0) \\
= \Phi(X'\beta / \sigma) \left( X'\beta + \frac{\sigma \phi(X'\beta / \sigma)}{\Phi(X'\beta / \sigma)} \right) \\
= \Phi(X'\beta / \sigma) X'\beta + \sigma \phi(X'\beta / \sigma).
\]

We can get the marginal effects of \( x_k \) on \( y \) by simply taking the first derivative of \( E[y \mid X] \) with respect to \( x_k \):
\[
\frac{\partial E[y \mid X]}{\partial x_k} = \frac{\partial \Pr(y > 0 \mid X)}{\partial x_k} E(y \mid X, y > 0) + \Pr(y > 0 \mid X) \frac{\partial E(y \mid X, y > 0)}{\partial x_k}
\]
The first term indicates the marginal changes in the probability of $y_i$ being positive times the expected value of $y_i$ conditioning on $x_i$. The second term indicates the marginal changes in the expected value of $y_i$ given the probability of $y_i$ being positive. This turns out to be very simple:

$$\frac{\partial E[y \mid X]}{\partial x_k} = \phi(X'\beta / \sigma)(\beta_k / \sigma)X'\beta + \Phi(X'\beta / \sigma)\beta_k - \sigma(X'\beta_i / \sigma)\phi(X'\beta_i / \sigma)(\beta_k / \sigma)$$

$$= \Phi(X'\beta / \sigma)\beta_k$$

Thus, the marginal effect of $x_k$ is a product of the probability of $y$ being larger than zero, $\Phi(X'\beta / \sigma)$, and the coefficient of $x_k$. Because the probability is between zero and one, the probability is like a scaling factor which scales down the size of the coefficient of $x_i$. If $y_i$ is never censored, then the probability of $y_i$ being larger than zero is one and the coefficient of $x_i$ become the marginal effect $x_i$.

If we only want to find the marginal effect of $x_k$ on $y$ when $y$ is larger than zero, we take the first derivative of $E[y \mid x, y > 0]$:

$$\frac{\partial E[y \mid X, y > 0]}{\partial x_i} = \frac{\partial(X'\beta + \lambda(X\beta / \sigma))}{\partial x_i}$$

$$= \beta_k + \sigma - \frac{(X'\beta / \sigma)\phi(X'\beta / \sigma)\Phi(X'\beta / \sigma)(\beta_k / \sigma) - \phi(X'\beta / \sigma)^2(\beta_k / \sigma)}{\Phi(\beta X_i / \sigma)^2}$$

$$= \beta_k - \beta_k (X'\beta / \sigma)\lambda(X'\beta / \sigma) - \lambda(X'\beta / \sigma)^2$$

$$= \beta_k [1 - \lambda(X'\beta / \sigma)((X'\beta / \sigma) + \lambda(X'\beta / \sigma))]$$

$$= \beta_k \Theta$$

Again, the coefficient of $x_i$ has a scaling factor, $\Theta$. This scaling factor is also between zero and one.

The difference between the marginal effects of $x_k$ on $y$ and $y > 0$ is that the former marginal effects include a possibility that an increase in $x_i$ may bring up some observations where $y_i = 0$ into positive, while the later marginal effects do not include such changes in the probability.

The likelihood of Tobit model is

$$L = \prod_{y_i=0} [1 - \Phi(X'\beta / \sigma)] \prod_{y_i>0} \sigma^{-1} \phi(y - X'\beta / \sigma)$$

The log likelihood is

$$\ln L = 1(y_i > 0) \ln [1 - \Phi(X'\beta / \sigma)] + 1(y_i > 0) \ln [\phi(y - X'\beta / \sigma)] - \ln(y_i > 0) \ln(\sigma)$$
Lecture 14B: The Sample Selection (Incidental Selection) Model

Based on the discussion above,

\[ y_i = \beta' x_i + u_i \quad u_i \sim N[0, \Phi^2] \]

\[ E[y_i | y_i > a] = \beta' x + \sigma \lambda(\alpha_y) \]

Suppose that \( y \) and \( z \) have a bivariate distribution with correlation \( \Delta \). We are interested in the distribution of \( y \) given that \( z \) exceeds a particular value.

\[ f(y, z | z > a) = \frac{f(y, z)}{Pr(z > a)} \]

\[ E[y | z > a] = \mu_y + \rho \sigma_y \lambda(\alpha_z) \]

Selection Mechanism

\[ z_i^* = \gamma' w_i + v_i \]

\[ z_i = 1 \quad \text{if} \quad z_i^* > 0, \]

\[ z_i = 0 \quad \text{if} \quad z_i^* <= 0, \]

\[ Pr(z_i = 1) = \Phi(\gamma' w_i) \]

\[ Pr(z_i = 0) = 1 - \Phi(\gamma' w_i) \]

Regression Model

\[ y_i = \beta' x_i + u_i \quad \text{if} \quad z_i = 1 \]

\[ E[y_i | z_i = 1] = \beta' x + \rho \sigma \lambda(\gamma' w) \]

The Heckman’s two-step estimation:

1. Estimate the probit equation to obtain (and compute
\[ \hat{\lambda}_i = \frac{\phi(\hat{\gamma}_i \omega_i)}{\Phi(\hat{\gamma}_i \omega_i)} \]

2. Estimate \( \beta' \) by including \( \hat{\lambda}_i \) in OLS of \( y \) on \( x \).

Note that,

1. \( \omega_i \) should be a sub set of \( x_i \)

2. at least one variable of \( \omega_i \) should not be in \( x_i \) (identification).

It turns out that the second requirement is often difficult and controversial.

**Example 14-2: Women’s labor supply, MROZ.dta**

See Example 17.5 in Wooldridge (1999).

**OLS**

\[ . \text{reg lwage educ exper expersq} \]

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 428</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model</td>
<td>35.0222967</td>
<td>3</td>
<td>11.6740989</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>188.305144</td>
<td>424</td>
<td>0.444115906</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>223.327441</td>
<td>427</td>
<td>0.523015084</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adj R-squared = 0.1509</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Root MSE = 0.66642</td>
</tr>
</tbody>
</table>

| lwage | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] |
|--------|-------|-----------|---|------|---------------------|
| educ | .1074896 | .0141465 | 7.60 | 0.000 | .0796837 .1352956 |
| exper | .0415665 | .0131752 | 3.15 | 0.002 | .0156697 .0674633 |
| expersq | -.0008112 | .0003932 | -2.06 | 0.040 | -.0015841 -.0000382 |
| _cons | -.5220406 | .1986321 | -2.63 | 0.009 | -.9124667 -.1316144 |
Heckman's Two-Step Estimation

. heckman lwage educ exper expersq, sel(educ exper expersq age kidslt6 kidsge6 nwifeinc) two

Heckman selection model -- two-step estimates
(regression model with sample selection)

Number of obs = 753
Censored obs = 325
Uncensored obs = 428

Wald chi2(6) = 180.10
Prob > chi2 = 0.0000

------------------------------------------------------------------------------
|      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lwage        |
    educ |   .1090655    .015523     7.03   0.000     .0786411      .13949
    exper |   .0438873   .0162611     2.70   0.007     .0120163    .0757584
    expersq |  -.0008591   .0004389    -1.96   0.050    -.0017194    1.15e-06
    _cons |  -.5781032   .3050062    -1.90   0.058    -.1175904     .019698
-------------+----------------------------------------------------------------
select       |
    educ |   .1309047   .0252542     5.18   0.000     .0814074     .180402
    exper |   .1233476   .0187164     6.59   0.000     .0866641    .1600311
    expersq |  -.0018871      .0006    -3.15   0.002     -.003063   -.0007111
    age |  -.0528527   .0084772    -6.23   0.000    -.1006288   -.0050766
    kidslt6 |  -.8683285  1185223    -7.33   0.000    -1.006282    -.769663
    kidsge6 |   .036005    .0434768     0.83   0.408     -.049208   .1212179
    nwifeinc |  -.0120237   .0048398    -2.48   0.013     -.0215096   -.0025378
    _cons |   .2700768    .508593     0.53   0.595    -.7267472    1.266901
-------------+----------------------------------------------------------------
mills        |
     lambda |   .0322619   .1336246     0.24   0.809    -.2296376    .2941614
-------------+----------------------------------------------------------------
          rho |    0.04861
Heckman’s Model with Maximum Likelihood Model

. heckman lwage educ exper expersq, sel(educ exper expersq age kidslt6 kidsge6 nwifeinc)

Iteration 0:   log likelihood = -832.89776
Iteration 1:   log likelihood = -832.88509
Iteration 2:   log likelihood = -832.88508

Heckman selection model                         Number of obs      =       753
(regression model with sample selection)        Censored obs       =       325
Uncensored obs     =       428

Wald chi2(3)       =     59.67
Log likelihood = -832.8851                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
|      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lwage        |
  educ |   .1083502   .0148607     7.29   0.000     .0792238    .1374767
  exper |   .0428369   .0148785     2.88   0.004     .0136755    .0719983
  expersq |  -.0008374   .0004175    -2.01   0.045    -.0016556   -.0000192
   _cons |  -.5526973   .2603784    -2.12   0.034     -.106303   -.0423651

select       |
  educ |   .1313415   .0253823     5.17   0.000     .0815931    .1810899
  exper |   .1232818   .0187242     6.58   0.000     .0865831    .1599806
  expersq |  -.0018863   .0004175    -4.47   0.000     -.0036933   -.0000803
    age |  -.0528287   .0084792    -6.23   0.000     -.1654621    -.0401957
  kidslt6 |  -.8673988   .1186509    -7.31   0.000     -1.09995   -.6348472
  kidsge6 |   .0358723   .0434753     0.83   0.409     -.0493377    .1210824
  nwifeinc |  -.0121321   .0048767    -2.49   0.013    -.0216903    -.002574
------------------------------------------------------------------------------
Example 14-3: Women’s Labor Supply, hours

.tobit hours nwifeinc educ exper expersq age kidslt6 kidsge6, ll(0)

Tobit estimates  Number of obs =    753
LR chi2(7) =   271.59
Prob > chi2 =  0.0000
Log likelihood = -3819.0946  Pseudo R2 =   0.0343

|         Coef.     Std. Err.     t    P>|t|    [95% Conf. Interval] |
|------------------|-----------------|------|-----|-------------------------|
| hours            |                 |      |     |                         |
| nwifeinc         |  -8.814243      | 4.459096 | -1.98 | 0.048       | -17.56811 | -.0603725 |
| educ             |   80.64561      | 21.58322 |  3.74 | 0.000       |  38.27453 | 123.0167  |
| exper            |  131.5643       | 17.27938 |  7.61 | 0.000       |   97.64231 | 165.4863  |
| expersq          |  -1.864158      | 0.5376615 | -3.47 | 0.001       |  -2.919667 |  -0.808679 |
| age              |  -54.40501      | 7.418496 | -7.33 | 0.000       |  -88.96862 |  -39.8414 |
| kidslt6          |  -894.0217      | 111.8779 | -7.99 | 0.000       |  -1113.655 |  -674.3887 |
| kidsge6          |   -16.218       | 38.64136 | -0.42 | 0.675       |   -92.07675 |   59.64075 |
| _cons            |   965.3053      | 446.4358 |  2.16 | 0.031       |   88.88531 | 1841.725  |

|         Coef.     Std. Err.     t    P>|t|    [95% Conf. Interval] |
|------------------|-----------------|------|-----|-------------------------|
| hours            |                 |      |     |                         |
| _se              |  1122.022       | 41.57903 |     |                         |

(Ancillary parameter)

Obs. summary: 325 left-censored observations at hours<=0
428 uncensored observations
Lecture 15: Attrition in Panel Data

Attrition in Panel Data

Motivation
In panel data, it is often the case where observations (countries, states, households, individuals, etc.) drop out of sample after \( t = 2 \) (Attrition). If the attrition happens randomly, then attrition does not create any serious problems. We would simply have a smaller number of observations. However, if the attrition takes place systematically, then the attrition may create a sample selection bias.

Suppose that we are interested in estimating the first difference model to eliminate the fixed effects:

\[
\Delta y_t
\]

\[
\Delta x_t
\]
\[ \Delta y_{it} = \Delta x_{it} \beta + \Delta u_{it} \] (1)

But suppose that samples identified in the circle in Figure 1 are not available at \( t = 2 \). Because the samples lost due to attrition have low values in both \( \Delta y_{it} \) and \( \Delta x_{it} \), the estimated coefficient of \( \Delta x_{it} \) would be biased downward.

The attrition selection model is

\[ s_{it} = I[w_{it} \hat{\delta}_t + v_{it} > 0] \] (2)

\( s_{it} \) is a dummy variable which is one if a sample is dropped out at time \( t \). Since we only consider absorbing states, \( s_{it} \) remains one if a sample drops out. \( w_{it} \) is a vector of variables that affect the attrition. \( w_{it} \) must be available for all observations.

Because the attrition problem can be considered as a sample selection problem, we can correct the sample selection by using the inverse Mills ratio, provided we can identify the selection with some identifying variables.

From the selection model, we can estimate the probability of each sample to be dropped out in each time period and use the predicted probability to create the inverse Mills ratio as

\[ \hat{\lambda}(w_{it} \hat{\delta}_t) = \phi(w_{it} \hat{\delta}_t) / \Phi(w_{it} \hat{\delta}_t) \] (3)

Thus we can estimate

\[ \Delta y_{it} = \Delta x_{it} \beta + \rho_2 d2 \hat{\lambda}_{it2} + \cdots + \rho_T dT \hat{\lambda}_{iT} + \text{error}_{it} \] (4)

where \( d2, \ldots, dT \) are time dummies.

**Inverse Probability Weighting (IPW) Approach**

Another method is based on inverse probability weighting (IPW). This method relies on an assumption that the observed information in the first period, \( z_{i1} \), is a strong predictor of selection in each time period (**selection on observables**). Notice that the information in the first period is observable for all samples. So, we have
Inverse Probability Weighting Approach

Step 1 Estimate the selection model (2) and obtain predicted probability,
Step 2 Use $1/\hat{p}_{it}$ as the weight in the regression models.

For the observations after $t = 3$, the weight should be calculated by multiplying the probabilities: $(1/\hat{p}_{i2})(1/\hat{p}_{i3})\ldots1/\hat{p}_{iT}$. 

\[
P(s_{it} = 1 \mid y_{it}, x_{it}, z_{it}) = P(s_{it} = 1 \mid z_{it}) \quad t = 2, \ldots, T
\]

Table 1 shows an example of attrition in panel data. Of the 1,500 households in the original data, 1,422 households (94.8 percent) were re-interviewed in 2000, and of those, 1,266 households (84.4 percent) were re-interviewed in 2002.

In Table 1, households that were interviewed in the first round, not in the second round, but again re-interviewed in the third round are not included. All the samples that were not interviewed in the second round are not included in Table 1. This is called “attrition is in an absorbing state.” In this lecture note, we only discuss attrition in an absorbing state.

Table 1. Sample Households and Working-Age Adult Mortality in Rural Kenya

<table>
<thead>
<tr>
<th>Province</th>
<th>Original sample households in 1997</th>
<th>Percentage of households re-interviewed among the original (1997) households</th>
<th>Percentage of households incurring working-age mortality between 1997-2000</th>
<th>Percentage of households incurring working-age mortality between 2000-2002</th>
<th>HIV prevalence at urban sentinel sites in 1990-91</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(A)</td>
<td>(B)</td>
<td>(C)</td>
<td>(D)</td>
<td>(E)</td>
</tr>
<tr>
<td>Coastal</td>
<td>113</td>
<td>- (96.7)</td>
<td>7 (6.0)</td>
<td>6.8</td>
<td>5.6</td>
</tr>
<tr>
<td>Eastern</td>
<td>242</td>
<td>(86.3)</td>
<td>(88.8)</td>
<td>4.3</td>
<td>2.7</td>
</tr>
<tr>
<td>Nyanza</td>
<td>280</td>
<td>(93.6)</td>
<td>(87.5)</td>
<td>12.2</td>
<td>3.3</td>
</tr>
<tr>
<td>Western</td>
<td>305</td>
<td>(95.7)</td>
<td>(89.4)</td>
<td>1.8</td>
<td>2.0</td>
</tr>
<tr>
<td>Central</td>
<td>181</td>
<td>(96.1)</td>
<td>(90.6)</td>
<td>2.9</td>
<td>2.4</td>
</tr>
<tr>
<td>Rift Valley</td>
<td>103</td>
<td>(93.1)</td>
<td>(71.2)</td>
<td>0.1</td>
<td>1.7</td>
</tr>
<tr>
<td>Total</td>
<td>1,500</td>
<td>1,122</td>
<td>1,200</td>
<td>7.8</td>
<td>2.3</td>
</tr>
</tbody>
</table>


*Note:* (a) Working-age is defined as 15-19 for women and 15-54 for men. (b) The average percentage of pregnant women who visited the urban sentinel-surveillance sites and tested HIV positive in 1990-1994. Data are taken from 11 urban sentinel-surveillance sites (NASCOP, 2001).
### Table 4. Household-level re-interview model (Probit*)

<table>
<thead>
<tr>
<th>Household &amp; Sectoral Rate</th>
<th>POOL</th>
<th>Attrited in 2000, continued with 1997 sample</th>
<th>Attrited in 2002 survey, continued with 1997 and 2000 sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(A)</td>
<td>(B)</td>
<td>(C)</td>
</tr>
<tr>
<td><strong>Longest HH Prevalence Rate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIV* In-Household Rate</td>
<td>0.071</td>
<td>0.002</td>
<td>0.529</td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td>(0.071)</td>
<td>(1.98)**</td>
</tr>
<tr>
<td><strong>Household Characteristics in 1997</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male: Some Primary School (−1)</td>
<td>0.008</td>
<td>-0.030</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
<td>(0.62)</td>
<td>(0.84)</td>
</tr>
<tr>
<td>Female: Some Primary School (−1)</td>
<td>0.018</td>
<td>0.028</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>(0.94)</td>
<td>(1.04)</td>
<td>(0.48)</td>
</tr>
<tr>
<td>Female: Primary Finished (−1)</td>
<td>0.007</td>
<td>0.009</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>(0.91)</td>
<td>(0.57)</td>
<td>(1.16)</td>
</tr>
<tr>
<td>Female: Primary Finished (−1)</td>
<td>0.002</td>
<td>0.012</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>(0.94)</td>
<td>(0.99)</td>
<td>(0.94)</td>
</tr>
<tr>
<td>Female Headed in 1997 (−1)</td>
<td>0.033</td>
<td>0.023</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>(11.15)</td>
<td>(12.22)</td>
<td>(0.88)</td>
</tr>
<tr>
<td>Polygamous Household in 1997 (−1)</td>
<td>0.060</td>
<td>0.071</td>
<td>N.A.</td>
</tr>
<tr>
<td>Number of Male Adults in 1997</td>
<td>0.002</td>
<td>0.002</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>(1.58)</td>
<td>(0.39)</td>
<td>(1.84)</td>
</tr>
<tr>
<td>Number of Female Adults in 1997</td>
<td>-0.005</td>
<td>-0.002</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>(1.15)</td>
<td>(0.39)</td>
<td>(0.68)</td>
</tr>
<tr>
<td>Land Tenure in 1997 (−1)</td>
<td>-0.081</td>
<td>-0.005</td>
<td>0.003</td>
</tr>
<tr>
<td>In (Landholding Size in Acres in 1997)</td>
<td>0.005</td>
<td>0.007</td>
<td>0.021</td>
</tr>
<tr>
<td>In (Asset Value in 1997)</td>
<td>0.001</td>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>Distance to Due Step in 1997 (km)</td>
<td>0.001</td>
<td>0.002</td>
<td>0.009</td>
</tr>
<tr>
<td>Distance to Piped Water in 1997 (km)</td>
<td>0.001</td>
<td>-0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>Year 2002 (−1)</td>
<td>-0.117</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>Econometric Terms Dummies</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Texn 2 in 2000</td>
<td>0.028</td>
<td>0.233</td>
<td>N.A.</td>
</tr>
<tr>
<td></td>
<td>(0.98)</td>
<td>(10.09)**</td>
<td>(0.16)</td>
</tr>
<tr>
<td>Texn 4 in 2000</td>
<td>-0.167</td>
<td>-0.059</td>
<td>N.A.</td>
</tr>
<tr>
<td></td>
<td>(1.76)</td>
<td>(1.85)</td>
<td>(0.84)</td>
</tr>
<tr>
<td>Texn 2 in 2002</td>
<td>0.133</td>
<td>N.A.</td>
<td>0.164</td>
</tr>
<tr>
<td></td>
<td>(4.40)**</td>
<td>(1.38)</td>
<td>(5.59)**</td>
</tr>
<tr>
<td>Texn 4 in 2002</td>
<td>-0.109</td>
<td>N.A.</td>
<td>-0.117</td>
</tr>
<tr>
<td></td>
<td>(0.83)</td>
<td>(2.79)**</td>
<td>(0.47)</td>
</tr>
<tr>
<td>Presence Dummies included**</td>
<td>Y.E.S.</td>
<td>Y.E.S.</td>
<td>Y.E.S.</td>
</tr>
<tr>
<td>Some Tests for Terms Effects (χ²)</td>
<td>39.1 [0.00]**</td>
<td>111.5 [0.00]**</td>
<td>24.6 [0.00]**</td>
</tr>
<tr>
<td>Joint Test for HOH Characteristics (χ²)</td>
<td>17.7 [0.34]</td>
<td>14.3 [0.04]</td>
<td>33.1 [0.04]**</td>
</tr>
<tr>
<td>E.F.D.</td>
<td>0.920</td>
<td>0.548</td>
<td>0.890</td>
</tr>
</tbody>
</table>

**Notes:** Numbers in parentheses are absolute z-scores, calculated using heteroskedasticity robust standard errors clustered for households. ** indicates 1 percent significance level; * indicates 5 percent significance level. (a) Estimated coefficients are marginal changes in probability. (b) Because all the polygamous households were re-interviewed in 2002, the dummy is excluded in column C. (c) Five province dummies are included but not reported in this table.
Table 5. Factors Associated with Households Afflicted by Working-age Adult Mortality (Probit)  

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(A)</td>
<td>Corrected for selection</td>
<td>Corrected for selection</td>
</tr>
<tr>
<td></td>
<td>(B)</td>
<td>(C)</td>
<td>(D)</td>
</tr>
<tr>
<td>Logged HIV Prevalence rate</td>
<td>0.761</td>
<td>0.369</td>
<td>0.516</td>
</tr>
<tr>
<td>Ratios of HIV - Pregnant Women</td>
<td></td>
<td></td>
<td>(3.61)**</td>
</tr>
<tr>
<td>Household Level Variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male: Some Primary (-1)</td>
<td>0.010</td>
<td>0.010</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.031)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>Male: Primary Finished (-1)</td>
<td>-0.011</td>
<td>-0.011</td>
<td>-0.011</td>
</tr>
<tr>
<td></td>
<td>(0.077)</td>
<td>(0.077)</td>
<td>(0.077)</td>
</tr>
<tr>
<td>Female: Some Primary (-1)</td>
<td>0.004</td>
<td>0.006</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(0.237)</td>
<td>(0.244)</td>
<td>(0.244)</td>
</tr>
<tr>
<td>Female: Primary Finished (-1)</td>
<td>-0.009</td>
<td>-0.009</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>(0.436)</td>
<td>(0.436)</td>
<td>(0.436)</td>
</tr>
<tr>
<td>Female Headed in 1997 (-1)</td>
<td>0.003</td>
<td>0.003</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>(0.131)</td>
<td>(0.131)</td>
<td>(0.131)</td>
</tr>
<tr>
<td>Polygamous III in 1997 (-1)</td>
<td>0.014</td>
<td>0.015</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>(1.54)</td>
<td>(1.60)</td>
<td>(1.84)</td>
</tr>
<tr>
<td># of Male Adults in 1997</td>
<td>0.005</td>
<td>0.005</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>(1.95)</td>
<td>(1.95)</td>
<td>(1.95)</td>
</tr>
<tr>
<td># of Female Adults in 1997</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.14)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Land Tenure in 1997 (-2)</td>
<td>0.010</td>
<td>0.010</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(2.12)**</td>
<td>(2.05)**</td>
<td>(1.17)</td>
</tr>
<tr>
<td>Ln (Landholding in 1997)</td>
<td>0.065</td>
<td>0.066</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td>(1.21)</td>
<td>(1.29)</td>
<td>(1.29)</td>
</tr>
<tr>
<td>Ln (Asset Values in 1997)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.12)</td>
<td>(0.12)</td>
</tr>
<tr>
<td>Distance to Road in 1997</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>(0.71)</td>
<td>(0.71)</td>
<td>(0.71)</td>
</tr>
<tr>
<td>Distance to Piped Water in 1997</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(1.07)</td>
<td>(1.07)</td>
<td>(1.07)</td>
</tr>
<tr>
<td>Year 2002 (-1)</td>
<td>0.033</td>
<td>0.032</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>(4.58)**</td>
<td>(4.50)**</td>
<td>(4.50)**</td>
</tr>
<tr>
<td>Province Dummies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nyanza province</td>
<td>0.036</td>
<td>0.034</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>(1.37)</td>
<td>(1.37)</td>
<td>(1.37)</td>
</tr>
<tr>
<td>Central province</td>
<td>0.016</td>
<td>0.019</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>(1.10)</td>
<td>(1.10)</td>
<td>(1.10)</td>
</tr>
<tr>
<td>Western province</td>
<td>0.001</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.10)</td>
<td>(0.10)</td>
</tr>
<tr>
<td>Central province</td>
<td>0.001</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.75)</td>
<td>(0.75)</td>
<td>(0.75)</td>
</tr>
<tr>
<td>Rift Valley province</td>
<td>0.008</td>
<td>0.008</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>(0.77)</td>
<td>(0.77)</td>
<td>(0.77)</td>
</tr>
</tbody>
</table>

Note: Numbers in parentheses are absolute t-scores, calculated using heteroskedasticity robust standard errors clustered for households. ** indicates 1 percent significance level, * indicates 5 percent significance level.  
(a) Estimated coefficients are marginal changes in probability.