Graduate Econometrics
Lecture 6
Models with Limited Dependent Variables

Yonas Alem

Department of Economics
University of Gothenburg

December 12, 2014

Maximum Likelihood Estimation Recap
Introduction

- ML method assumes knowledge of the entire distribution, not just of a number of its moments as in GMM method
- If these distributional assumptions are correct, the ML estimator, is under weak regularity conditions, consistent and asymptotically normal.
- It is also asymptotically efficient since it fully exploits the assumptions about the distribution.
- Starting assumptions of ML:
  - The conditional distribution of an observed phenomenon is known, except for a finite number of unknown parameters.
  - These parameters will be estimated by taking those values for them that give the observed values the highest probability, the highest likelihood.
  - It provides an approach of estimating a set of parameters characterizing a distribution, if we know, or assume we know, the form of this distribution

Maximum Likelihood Estimation Recap
Introduction

Example -1

- Consider a large pool of balls filled with red and yellow balls
- One could be interested in the fraction p of red balls in this pool
- Take random sample of N balls only
- Let $y_i = 1$ if ball i is red and $y_i = 0$ otherwise
- Thus, $P\{y_i = 1\} = p$
- Suppose the pool of ball contains $N_1 = \sum_{i=1}^{N} y_i$ red and $N - N_1$ yellow balls

Maximum Likelihood Estimation Recap

Introduction Cont.

Example -1 Cont.

- The likelihood (probability) of obtaining such a sample is given by:

$$P\{N_1 \quad red \quad balls, \quad N - N_1 \quad yellow \quad balls\} = p^{N_1}(1-p)^{N-N_1} \tag{1}$$

- Equation 1 is what is called the **Likelihood Function** and it is a function of the unknown parameter p.
- In ML estimation, we choose a value for p such that the likelihood function is maximal, and obtain $\hat{p}$.
- The conventional practice is to maximize the log-likelihood, which is a simple monotonic transformation of equation [1] (for computational convenience)

$$logL(p) = N_1 log(p) + (N - N_1) log(1-p) \tag{2}$$

Maximum Likelihood Estimation Recap

Introduction Cont.

Example -1 Cont.

- FOC to maximize [1]:

$$\frac{dlogL(p)}{dp} = \frac{N_1}{p} - \frac{N - N_1}{1 - p} = 0 \quad (3)$$

- Solving [3] for p gives the ML estimator $\hat{p} = N_1 / N$
- It corresponds to the sample proportion of red balls, and most likely to your best guess for p based on the sample drawn
- SOC:

$$\frac{d^2 logL(p)}{dp^2} = \frac{N_1}{p^2} - \frac{N - N_1}{(1 - p)^2} < 0 \quad (4)$$

- Indicating that we indeed have a maximum
- Another example:

Maximum Likelihood Estimation Recap

Introduction Cont.

Example 2.

- Consider the simple regression model

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \tag{5}$$

- Keep assumptions [A1][A4]
- The assumptions imply that $E\{y_i|x_i\} = \beta_1 + \beta_s x_i$ & $V\{y_i|x_i\} = \sigma^2$
- To estimate the above model, we need to impose distributional assumption on ε, the most common being assumption [A5] (normal dist.)

Maximum Likelihood Estimation Recap

Introduction Cont.

Example 2.

- The contribution of the i^{th} observation to the likelihood function is the value of the density function at the observed point y_i. Which, for a normal distribution yields,

$$f(y_i|x_i;\beta,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\{-\frac{1}{2}\frac{(y_i-\beta_1-\beta_2 x_i)^2}{\sigma^2}\} \quad (6)$$

- **Note**: y_i has a continues distribution, hence the likelihood of observing a particular outcome y for y_i is zero for any y
- Where $\beta = (\beta_1, \beta_2)$
- The joint density of $(y_1, ..., y_N)$ conditional on $X = (x_1, ..., x_N)'$ is stated as

$$f(y_1,...,y_N|X;\beta,\sigma^2) = \prod_{i=1}^{N} f(y_i|x_i;\beta,\sigma^2) = (\frac{1}{\sqrt{2\pi\sigma^2}})^N \prod_{i=1}^{N} exp\{-\frac{1}{2}\frac{(y_i -$$

(7)

Maximum Likelihood Estimation Recap

Introduction Cont.

Example 2. Cont.

- The likelihood function and the joint density function of $y_1, ..., y_N$ are similar except the fact that the former is considered as a function of the unknown parameters β, σ^2
- The LL function is given by

$$logL(\beta, \sigma^2) = -\frac{N}{2} log(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{N}\frac{(y_i - \beta_1 - \beta_2 x_i)^2}{\sigma^2} \quad (8)$$

- Maximizing (8) w.r.t $\beta_1 \& \beta_2$ corresponds to minimizing the residual sum of squares $S(\beta)$, as shown in OLS. Do you see why?

Maximum Likelihood Estimation Recap

Introduction Cont.

Example 2. Cont.

- Meaning that the ML estimators of $\beta_1 \& \beta_2$ are identical to the OLS estimators!
- Denote these estimators by $\hat{\beta}_1$ and $\hat{\beta}_2$, and define the residuals $e_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i$ and maximize (8) w.r.t σ^2. FOC:

$$-\frac{N}{2}\frac{2\pi}{2\pi\sigma^2} + \frac{1}{2}\sum_{i=1}^{N}\frac{e_i^2}{\sigma^4} = 0 \tag{9}$$

- Solve(9) for σ^2 to get the ML estimator for σ^2 given by

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} e_i^2 \tag{10}$$

Maximum Likelihood Estimation Recap

Introduction Cont.

Example 2. Cont.

- Note that however this estimator is consistent but not unbiased (a small sample problem) as the estimator in OLS which was given by

$$s^2 = \frac{1}{N-K} \sum_{i=1}^{N} e_i^2 \tag{11}$$

- In many cases, the ML estimator cannot be shown to be unbiased (unknown small sample properties)
- Its use generally is defended based on asymptotic grounds
- Analytical solution of the ML estimator is also difficult in many cases except in some general cases as shown above

Maximum Likelihood Estimation Recap

Specification Tests

- Three types of tests
  1. The Wald test: pretty much in line with t and F tests
  2. The likelihood ratio test: used to compare two alternative nested models
  3. The lagrange multiplier test: used to test restrictions imposed in estimation

Cross-sectional Binary Choice Models

- Used to model phenomenal that are of discrete nature
  - Do married women participate in the labor force?
  - Which sections of society are poor?
  - What are the determinants of an agricultural technology adoption?
- For such kinds of models, OLS is generally inappropriate - we rather use binary choice models
- Mostly (although not exclusively) the problems analyzed are micro-economic nature

Cross-sectional Binary Choice Models Cont.

- Suppose we want to study the impact of income (assumed as the only variable here) on the probability of owning a car:

$$y_i = \beta_1 + \beta_2 x_{i2} + \varepsilon_i = x_i'\beta + \varepsilon_i \tag{12}$$

- Where, $y_i = 1$ if family i owns a car, 0 if family i does not own a car
- $x_i = (x_{i1}, x_{i2})'$
- The standard assumptions:

$$E\{\varepsilon_i|x_i\} = 0 \quad such \;\; that \quad E\{y_i|x_i\} = x_i'\beta \Longrightarrow \tag{13}$$

$$E\{y_i|x_i\} = 1.P\{y_i = 1|x_i\} + 0.P\{y_i = 0|x_i\} \tag{14}$$

$$= 1.P\{y_i = 1|x_i\} = x_i'\beta \tag{15}$$

Cross-sectional Binary Choice Models Cont.

- Thus, the linear model implies that $x_i'\beta$ is a probability and should therefore lie between 0& 1.
- This is only possible if the x_i values are bounded and if certain restrictions on β are satisfied.
  - Hard to achieve this in practice
- Another fundamental problem:
  - ε_i in equation 12 has a highly non-normal distribution and suffers from heteroscedasticity
  - Because y_i has only two possible outcomes, so does the error term for a given value of x_i
- The distribution of ε_i can be summarized as:

Cross-sectional Binary Choice Models Cont.

- $$P\{\varepsilon_i = -x_i'\beta | x_i\} = P\{y_i = 0 | x_i\} = 1 - x_i'\beta \qquad (16)$$

$$P\{\varepsilon_i = 1 - x_i'\beta | x_i\} = P\{y_i = 1 | x_i\} = x_i'\beta \qquad (17)$$

- This implies that the variance of the error term is not constant but dependent upon the explanatory variables

$$V\{\varepsilon_i | x_i\} = x_i'\beta(1 - x_i'\beta) \qquad (18)$$

Cross-sectional Binary Choice Models Cont.

- We therefore use binary choice models (or univariate dichotomous models)
- Describe the probability $y_i = 1$ directly (but derived from an underlying latent variable model (see next pages)
- The general formulation is:

$$P\{y_i = 1|x_i\} = G(x_i, \beta) \tag{19}$$

for some function $G(.)$

- Equation $[19]$ says that the probability of having $y_i = 1$ depends on x_i
- But, clearly, $G(.)$ should take on values in the interval $[0, 1]$ only
- Usually, we assume:

$$G(x_i, \beta) = F(x_i'\beta) \tag{20}$$

Cross-sectional Binary Choice Models Cont.

- Common choices of F are the standard normal distribution

$$F(x'\beta) = \Phi(x'\beta) = \int_{-\infty}^{x'\beta} \Phi(z)\,dz, \tag{21}$$

giving rise to the so-called **Probit Model**, and the standard logistic function given by:

$$L(x'\beta) = \frac{e^{x'\beta}}{(1 + e^{x'\beta})} \tag{22}$$

leading to the **Logit Model**

Cross-sectional Binary Choice Models Cont.

- A third option is a uniform distribution over the interval [0,1] with distribution function:

$$F(x'\beta) = 0, x'\beta < 0; \tag{23}$$

$$F(x'\beta) = x'\beta, 0 \leq x'\beta \leq 1; \tag{24}$$

$$F(x'\beta) = 1, x'\beta > 1. \tag{25}$$

- Leading to what is called the **Linear Probability Model** pretty similar with [12], except that the probabilities are set to 0 or 1 if $x_i'\beta$ exceeds the lower upper limit respectively.

Cross-sectional Binary Choice Models Cont.

- Probit and logit are more common on applied work.
- Both the standard normal and the standard logistic random variable have an expectation of zero, while the latter has a variance of $\pi^2/3$ instead of 1.
- Correcting for the scaling difference would give similar results
- Apart from their signs, the coefficients in these binary choice models are not easy to interpret directly
- One needs to compute the marginal effects of changes in the explanatory variables

Cross-sectional Binary Choice Models Cont.

- For a continuous explanatory variable, x_{ik}, say the marginal effect is defined as the partial derivative of the probability that y_i equals one.
- For the three models above, we obtain

$$\frac{\partial \Phi(x_i'\beta)}{\partial x_{ik}} = \phi(x_i'\beta)\beta_k; \tag{26}$$

$$\frac{\partial L(x_i'\beta)}{\partial x_{ik}} = \frac{e^{x_i'\beta}}{(1+e^{x_i'\beta})^2}\beta_k \tag{27}$$

$$\frac{\partial (x_i'\beta)}{\partial x_{ik}} = \beta_k; (or \;\; 0) \tag{28}$$

Cross-sectional Binary Choice Models Cont.

- MEs are typically computed for the "average" observation, replacing x_i in the previous expressions with the sample averages.
- For the logit model, re-write [19] as

$$log\frac{p_i}{1-p_i} = x_i'\beta, \tag{29}$$

where $p_i = P\{y_i = 1|x_i\}$ is the probability of observing outcome 1.

- The lhs expression is known as the "log odds ratio"
- For example, an odds ratio of 3 mean that the odds of $y_i = 1$ are 3 times those of $y_i = 0$
- The β coefficients therefore can be interpreted as describing the effect upon the odds ratio
- If $\beta_k = 0.1$, a one-unit increase of x_{ik} increases the odds ratio by about 10% (ceteris paribus)

Cross-sectional Binary Choice Models
Estimation

- Very often binary choice modes are derived from underlying behavioral model - following the latent model approach.

$$y^* = x_i'\beta + \epsilon_i \tag{30}$$

- y^* is referred to as the latent variable because it is unobserved
- Assume a probability model of working where a person chooses to work if the utility difference exceeds a certain threshold level
- Thus, one observes $y_i = 1$ (working) if and only if $y_i^* > 0$, and $y_i = 0$ (not working) otherwise.
- Hence,

$$P\{y_i = 1\} = P\{y_i^* > 0\} = P\{x_i'\beta + \epsilon_i > 0\} = P\{-\epsilon_i \leq x_i'\beta\} = F(x_i'\beta) \tag{31}$$

Cross-sectional Binary Choice Models
Estimation

- Where F denotes the distribution function of $-\epsilon_i$
- Thus, depending on the distributional assumptions of ϵ_i, one can use either of the binary choice models
- A normalization on the distribution of ϵ_i is needed as the scale of utility is not identified
- Thus, we fix the variance at a given value and use the Maximum Likelihood method

Cross-sectional Binary Choice Models

Estimation Cont.

- The likelihood contribution of observation i with $y_i = 1$ is given by $P\{y_i = 1|x_i\}$ as a function of β. We do the same for $y_i = 0$
- We can write the likelihood function to be maximized for the entire sample as

$$L(\beta) = \prod_{i=1}^{N} P\{y_i = 1|x_i; \beta\}^{y_i} P\{y_i = 0|x_i; \beta\}^{1-y_i} \tag{32}$$

and the corresponding log-likelihood function (which is convenient to work with) will be given by

$$logL(\beta) = \sum_{i=1}^{N} y_i logF(x_i'\beta) + \sum_{i=1}^{N} (1 - y_i) log(1 - F(x_i'\beta)). \tag{33}$$

Cross-sectional Binary Choice Models

Estimation Cont.

- Substitute the appropriate function for F to get an expression to be maximized w.r.t β
- The values of β and their interpretation depends on the functional form used
- FOC: Differentiate [33] w.r.t. β to get

$$\frac{\partial logL(\beta)}{\partial\beta} = \sum_{i=1}^{N}[\frac{y_i - F(x_i'\beta)}{F(x_i'\beta)(1 - F(x_i'\beta))}f(x_i'\beta)]x_i = 0 \quad (34)$$

- Where $f = F'$ is the first derivative of the distribution function (so it is the density function)
- The term in the squared bracket is called the "generalized residual" of the model

Cross-sectional Binary Choice Models
Estimation Cont.

- It equals $f(x'\beta)/F(x_i'\beta)$ if $(y_i = 1)$ and $-f(x_i'\beta)/(1 - F(x_i'\beta))$ when $(y_i = 0)$
- The FOC says that each explanatory variable should be orthogonal to the generalized residual (over the whole sample)
- This is comparable with the OLS FOCs stating that the least square residuals are orthogonal to each variable in x_i
- For the Logit model, one can simplify [34] to

$$\frac{\partial logL(\beta)}{\partial\beta} = \sum_{i=1}^{N}[y_i - \frac{exp(x_i'\beta)}{1 + exp(x_i'\beta}]x_i = 0 \qquad (35)$$

Cross-sectional Binary Choice Models
Estimation Cont.

- The solution to [35] is the ML estimator of $\hat{\beta}$. From this estimate one can estimate the probability that $y_i = 1$ for a given x_i as

$$\hat{p} = \frac{exp(x_i'\hat{\beta})}{1 + exp(x_i'\hat{\beta}}) \tag{36}$$

Cross-sectional Binary Choice Models
Estimation Cont.

- Consequently, the FOC for the logit model imply that:

$$\sum_{i=1}^{N} \hat{p}_i x_i = \sum_{i=1}^{N} y_i x_i \tag{37}$$

- Thus including the constant term in the regression, the sum of the estimated probabilities is equal to $\sum_i y_i$ or the number of observations in the sample for which $y_i = 1$ I.e., the predicted frequency is equal to the actual frequency
- Similarly, if x_i includes a dummy variable, say 1 for employed people and 0 for unemployed, then the predicted frequency will be equal to the actual frequency for each labor market status group
- Pretty much the same holds for the probit model.
- SOCs, will imply that the matrix of the second order derivatives is negative definite (with the assumption of no

Cross-sectional Binary Choice Models

Goodness-of-fit

- A goodness-of-fit measure is a summary statistic indicating the accuracy with which the model approximates the observed data, like R^2 in OLS
- In binary choice models, the dependent variable is qualitative and accuracy can be judged either in terms of the fit between the calculated probabilities and observed response frequencies or in terms of the model's ability to forecast observed responses
- Unlike the linear regression model, there is no single measure of goodness-of-fit for binary choice models
- Often, goodness-of-fit measures are based on comparison with a model that contains only a constant as explanatory variable

Cross-sectional Binary Choice Models
Goodness-of-fit Cont.

- Let $logL_1$ represent the maximum loglikelihood value of the model of interest and let $logL_0$ denote the maximum value of the loglikelihood function when all parameters, except the intercept are zero
- Obviously, $logL_1 \geq logL_0 \implies$ the larger the difference between the two likelihood values, the more the extended model adds to the very restrictive model
- A formal likelihood ratio test can be based on the difference between the two values
- One popular measure goodness-of-fit measure is given by

$$pseudo - R^2 = 1 - \frac{1}{1 + 2(logL_1 - logL_0)/N} \tag{38}$$

Cross-sectional Binary Choice Models

Goodness-of-fit Cont.

- An alternative measure is suggested by McFadden (1974)

$$McFadden\ R^2 = 1 - logL_1 \,/\, LogL_0 \tag{39}$$

also known as the likelihood ratio index

- $logL_0 < LogL_1 < 0 \Longrightarrow$ both measures take on values in the interval $[0, 1]$

Cross-sectional Binary Choice Models

Specification Tests

- ML estimators have the property of being consistent but there is one important condition to hold
  - The likelihood function has to be correctly specified, i.e., we have to be sure about the entire distribution that we impose on our data
- Deviations will cause inconsistent estimators
  - This typically arises when $Pr(y = 1)$ is misspecified as a function of x_i
- Such misspecification are motivated from the latent variable model and reflect heteroskedasticity or non-normality (in the probit case) of ε_i
- In addition, one would want to test for omitted variables without having to re-estimate the model
- The most convenient test is the **Lagrange Multiplier test**

Cross-sectional Binary Choice Models

Specification Tests

- If we want to test for J omitted variables in the model, a simple way of computing the LM statistics is obtained from the a regression of a vector of ones upon the $K + J$ variables and computing N times the uncentered R^2 (the R^2 without a constant term) of this auxiliary regression
  - Under H_0 : that the J omitted variables enter the model with zero coefficients, the test statistic is asymptotically Chi-squared distributed with J degrees of freedom
- Heteroskedasticity: unlike the linear model, heteroskedasticity of ε_i leads to inconsistent parameter estimates
  - This can be tested using an LM test similar with Breusch-Pagan test
  - In the case of the probit model, one can estimate a heteroskedasticity consistent model (command: hetprobit)

Cross-sectional Binary Choice Models

Illustration

- Let's see some application on binary choice models. Source: the book "An Introduction to Modern Econometrics Using Stata", by Christopher Baum, Stata Press, 2006.
- Data: 2000 women, 657 of which are not recorded as wage earners
- The dependent variable (work) is set to zero for the nonworking and to one for those reporting positive wages
- First run s simple probit and then a simple logit model of a decision to work as a function of the woman's age, marital status, number of children, and level of education

Cross-sectional Binary Choice Models

Illustration

```
. probit work age married children education, nolog // regression

Probit regression                               Number of obs     =       2000
                                                LR chi2(4)        =     478.32
                                                Prob > chi2       =     0.0000
Log likelihood = -1027.0616                     Pseudo R2         =     0.1889

------------------------------------------------------------------------------
        work |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0347211   .0042293     8.21   0.000     .0264318    .0430105
     married |   .4308575    .074208     5.81   0.000     .2854125    .5763025
    children |   .4473249   .0287417    15.56   0.000     .3909922    .5036576
   education |   .0583645   .0109742     5.32   0.000     .0368555    .0798735
       _cons |  -2.467365   .1925635   -12.81   0.000    -2.844782   -2.089948
------------------------------------------------------------------------------

.
. mfx //marginal effects

Marginal effects after probit
      y  = Pr(work) (predict)
         =  .71835948
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
     age |    .011721      .00142    8.25   0.000   .008935  .014507   36.208
married* |    .150478      .02641    5.70   0.000   .098716   .20224    .6705
children |   .1510059      .00922   16.38   0.000   .132939  .169073   1.6445
educat~n |   .0197024       .0037    5.32   0.000   .012442  .026963   13.084
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Cross-sectional Binary Choice Models

Illustration

```
. logit work age married children education, nolog // regression

Logistic regression                               Number of obs   =       2000
                                                  LR chi2(4)      =     476.62
                                                  Prob > chi2     =     0.0000
Log likelihood = -1027.9144                       Pseudo R2       =     0.1882

------------------------------------------------------------------------------
        work |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0579303    .007221     8.02   0.000     .0437773    .0720833
     married |   .7417775   .1264705     5.87   0.000     .4938998    .9896552
    children |   .7644882   .0515289    14.84   0.000     .6634935     .865483
   education |   .0982513   .0186522     5.27   0.000     .0616936     .134809
       _cons |  -4.159247   .3320401   -12.53   0.000    -4.810034   -3.508461
------------------------------------------------------------------------------

.
. mfx //marginal effects

Marginal effects after logit
      y  = Pr(work) (predict)
         =  .72678588
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
     age |   .0115031      .00142    8.08   0.000   .008713  .014293   36.208
married* |   .1545671      .02703    5.72   0.000   .101592  .207542    .6705
children |    .151803      .00938   16.19   0.000   .133425  .170181   1.6445
educat~n |   .0195096       .0037    5.27   0.000    .01226   .02676   13.084
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Cross-sectional Binary Choice Models

Illustration

- Which one to use? The CDF's underlying these models differ most in the tails, producing similar predicted probabilities for nonextreme values of $x\beta$
- You can not directly test one against the other though because the likelihood estimators of the two estimators are not nested
- The parameter estimates from the probit and logit models will differ because they are estimates of (β/σ_u)
- But we saw that the variance of the standard normal distribution is unity, the variance of the logistic distribution is $\pi^2/3 \implies$ reported logit coefficients are larger by a factor of about $\pi/\sqrt{3} = 1.814$
- But as you saw, the marginal effects are similar and that is what we want. The choice between the two is, mainly a matter of taste!

Multiresponse Models

Ordered Response Models

- The number of responses for the dependent variable may be greater than 2
- Consider the question: "What are your alternative means of transportation?", possible responses "car", "bus", "bicycle"
  - You can see that there is no meaningful way to order these responses
  - Such models are estimated using multinomial models
- Consider the survey question on happiness: "Overall, how happy did you feel yesterday?"
  - The responses could be "Very Unhappy", "Unhappy", "Neither Happy nor Unhappy", "Happy", "Very Happy"
- You can see that there is some ordering in the responses $\Longrightarrow$ you can use ordered response models for estimation

Multiresponse Models

Ordered Response Models cont.

- Consider the choice between M alternatives, numbered from 1 to M (1-5 in the case of the happiness question
- This model is also based with one underlying latent variable but with a different match from the latent variable y_i^* to the ordered one $(y_i = 1, 2, ..., M)$

$$y_i^* = x_i'\beta + \varepsilon_i \tag{40}$$

$$y_i = j \quad if \quad \gamma_{j-1} < y_i^* \leq \gamma_j \tag{41}$$

for unknown $\gamma_j s$ with $\gamma_0 = -\infty, \gamma_1 = 0$ and $\gamma_M = \infty$

Multiresponse Models

Ordered Response Models cont.

- Consequently, the probability that alternative j is chosen is the probability that the latent variable y_i^* is between two boundaries γ_{j-1} and γ_j
- Assuming that ε_i is i.i.d, standard normal results in the **ordered probit model**. The logistic distribution gives the **ordered logit model**
- Do you see what happens when $M = 2$?,
- Stata commands: ordered probit (oprobit), ordered logit (ologit)

Tobit Models

- Tobit Models: in some applications the dependent variable is continuous, but its range may be constrained, e.g., it may be zero for a substantial part of most of the observations but postitive (with many different outcomes) for the rest of the population
    - E.g., expenditures on durable goods, hours of work, the amount of foreign direct investment
- Such types of models can be estimated using the **tobit model**
- This model is very similar with the probit model