JEFFREY M. WOOLDRIDGE

Introductory Econometrics
A Modern Approach

SIXTH EDITION

Chapter 17

Limited Dependent Variable Models and Sample Selection Corrections

Agenda:

(1) Qualitative response models

(2) The linear probability model (LPM). Applications

(3) The LOGIT model. Estimation & interpretations

(4) The PROBIT models

(1) Qualitative response models

Dependent variable is qualitative.

The qualitative dependent variable can be:

- **Binary, or dichotomous, variable**
- **Polychotomous (multiple category)**

Difference between models with quantitative & qualitative dependent variables

Quantitative Dependent variable	Qualitative Dependent variable
Our objective is to estimate the **<u>mean value</u>** of Y given the values of regressors Dependent variable has continuous nature	Our objective is to find the **<u>probability of something happening</u>**, hence the qualitative response models are also known as probability models

Three approaches to the developing a probability model for binary response variable:

1. The Linear probability model
2. The Logit Model
3. The Probit model

Linear probability model (LPM).

Consider the model

$$Y_i = \beta_1 + \beta_2 X_i + e_{1i}$$

where X= Family income and Y= if the family owns a house and 0 if it does not own a house.

$$E(Y_i | X_i) = \beta_1 + \beta_2 X_i$$

Now if P = the probability that Y_i=1 and (1-P) that Y=0, then this fallows Bernoulli probability distribution.

The mean of binomial distribution is np and variance np(1-p)

Problems with LPM

- Non normality of the disturbances u_i. Like Y_i, the disturbance term also accepts only values, that is they fallow Bernoulli probability distribution. As the sample size increases then the distribution converges to normal distribution

- Hetroscedastic variance. Since the mean and variance of Bernoulli distribution are equal to p and p(1-p) respectively, the variance of error term is hetroscedastic.

In the presence of hetroscedasticity OLS estimates are unbiased but have not the minimum variance.

the way to deal with hetroscedasticity is to transform the model by dividing it through by

$$\sqrt{E(Y_i / X_i)[1 - E(Y_i / X_I)} = \sqrt{P_i(1 - P_i)} = \sqrt{w_i}$$

$$\frac{Y_i}{\sqrt{w_i}} = \frac{\beta_1}{\sqrt{w_i}} + \beta_2 \frac{X_i}{\sqrt{w_i}} + \frac{u_i}{\sqrt{w_i}}$$

How to estimate w_i

- In practice E(Y/X) is not known thus fallowing procedure is used for finding *w.*

1. *1 Step.* Run OLS regression and obtain estimate of Y. then obtain $w = Y^{est}(1 - Y^{est})$
2. *2 Step* Use w estimated to transform the equation and estimated transformed model by OLS

Problems with LPM

- Non fulfillment of $0 <= E(Y_i | X) <= 1$

the probability of Y occurring given X should be between 1 and 0.

if it is not there are two ways to deal

1. assumptions

2.use Logit or Probit model

- Questionable R squared. Usually low R^2 Practical value of R^2 will be between 0.2 and 0.6

Problems with LPM

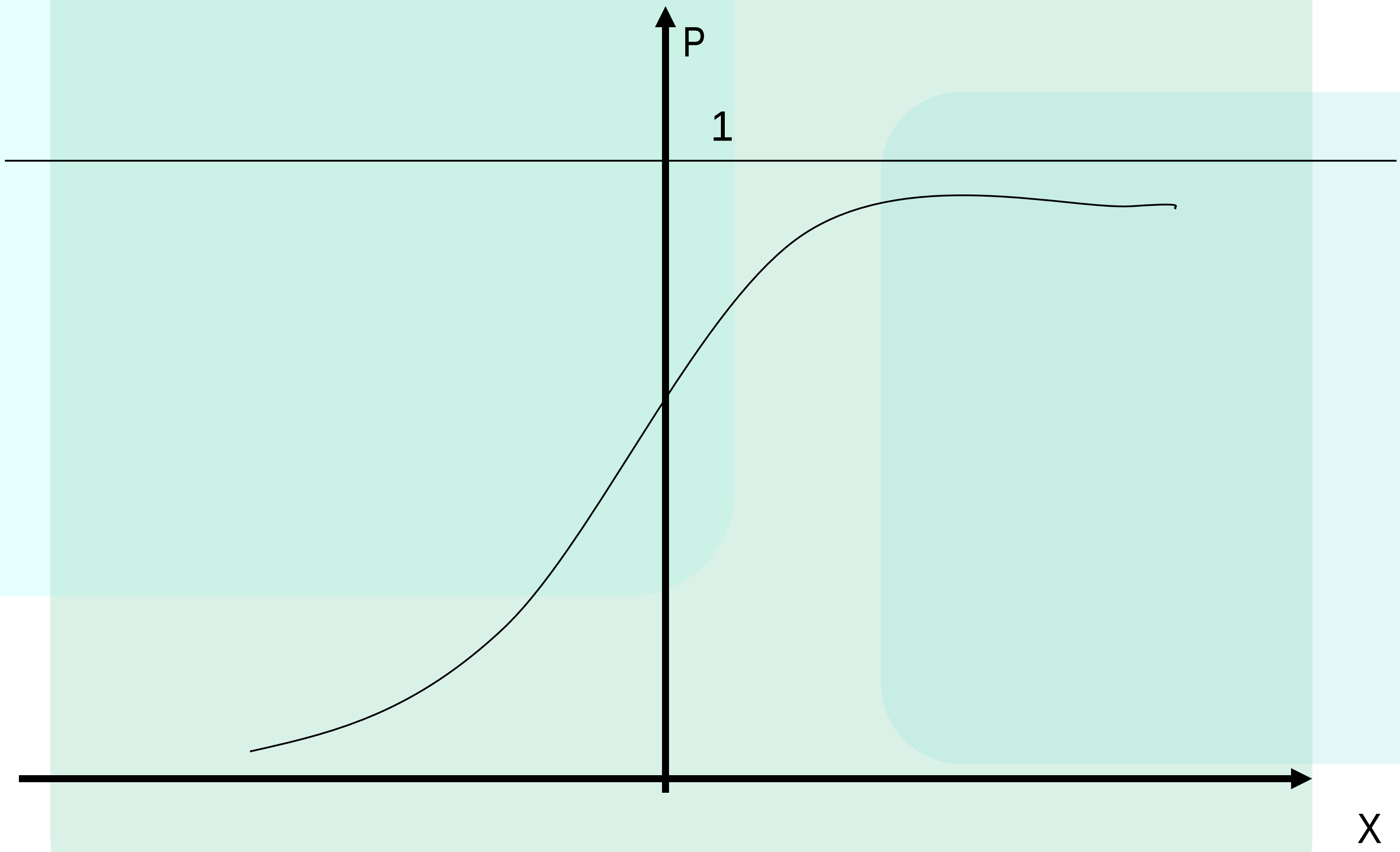
- The model is not logically very attractive since it assumes that P_i increases with X, that is marginal effect of X remains constant
- Thus we need model which has two features:

a) As X_i increases $P_i(Y_i / X_i)$ also increases but does not step outside 0 and 1.

b) The relationship between X and P is nonlinear, that is *S* shaped realtionship

These two features are possessed by Logit and Probit models

S shaped relationship



Logit model

Consider fallowing representation of home ownership

$$P_i = E(Y=1|X_i) = \frac{1}{1+e^{-(\beta_1+\beta_2 X_i)}}$$

or alternatively we can write the formula as

$$P_i = E(Y=1|X_i) = \frac{1}{1+e^{-Z_i}}$$

Which is also known as logistic distribution function, where $Z=\beta_1+\beta_2 X_i$

Z = [-∞:+∞]   P = [0 : 1]   P and Z are nonlinear. Which means it is impossible to use OLS without linearization.

P_i = probability of owning a house, 1-P_i probability of not owning a house

1-P=1/(1+e-^{Z}) therefore we can write

$$\frac{P_i}{1-P_i}=\frac{\dfrac{1}{1+e^{-Z_i}}}{\dfrac{e^{-Z_i}}{1+e^{-Z_i}}}=\frac{1}{e^{-Z_i}}=e^{Z_i}$$

Odds ratio

Features of Logit model

1. P_i = [0 : 1] Z_i = [-∞:+∞] L_i = [-∞:+∞]
2. L_i is linear in X_i, but P_i are not
3. Several regressors can be introduced into model
4. If logit is positive, then when the value of regressors increase, the odds that the regressand will equal to 1
5. The slope coefficient β_2 measure the change in L for a unit change in X,
6. Once we estimate the slope and intercept then we can estimate the probability of owning a house itself

Estimation of Logit model

We have to estimate fallowing models:

$$L_i = \ln(P_i / 1 - P_i) = Z_i = \beta_1 + \beta_2 X_i + u_i$$

For estimation we need values of X and L. The values of L depend on type of the data.

Data can be: a) Data at individual or micro level

b) grouped or replicated data

Grouped or replicated data

Consider the case of the table where the families are grouped according the income level tabel 15.4.

$$P^{est} = n_i / N_i$$

Which is relative frequency, if N is fairly large then the precision of the estimate will be better

$$L_i = \ln(P_i / 1 - P_i) = \beta_1 + \beta_2 X_i$$

Here the error term u, has fallowing feature

$$u_i \sim N\ [0,\ 1 / (N_i * P_i * (1 - P_i))]$$

Therefore the problem of hetroscedasticity is present thus we use WLS instead of OLS

Steps in estimating Logit model

- For each income X level estimate P=n/N
- For each X obtain Logit

$$L_i=\ln(P_i/1-P_i)$$

- Resolve the problem of hetroscedasticity

$$\frac{Y_i}{\sqrt{w_i}}=\frac{\beta_1}{\sqrt{w_i}}+\beta_2\frac{X_i}{\sqrt{w_i}}+\frac{u_i}{\sqrt{w_i}}$$

- Use OLS to estimate the above representation
- The conclusions are valid only if sample is reasonably large since the estimated equation fallow the regression through the origin

Interpretation

- $L = -1.59474\ w^{1/2} + 0.07862X_i$

se= (0.11046)   (0.00539)

t = (-14.43619)   (14.56675) $R^2=0.96$

- Logit interpretation, for a unit increase in weighted income, weighted log odds in favor of owning a house goes up by 0.08 units
- Odds interpretation, antilog of j-th coefficient, subtract 1 from it and , and multiply the result by 100, you will get the percentage change in the odds for a unit increase in j-th coefficient

Statistical inference and goodness of fit in the case of individual data

- Estimated standard errors are asymptotic
- Use Z statistic instead of t statistic

$$R^2 = \frac{\text{number of correct predictions}}{\text{total number of observations}}$$

in binary regression models the goodness of fit of secondary importance, the signs of coefficients and the statistical significance is of the first importance

- We use LR (Chı sqaure) test as a F test.

Probit model

- **Sometimes the logit model can use normal CDF, this kind of models are called probit or normit models:**

$$I = X\beta$$

Utility Index:

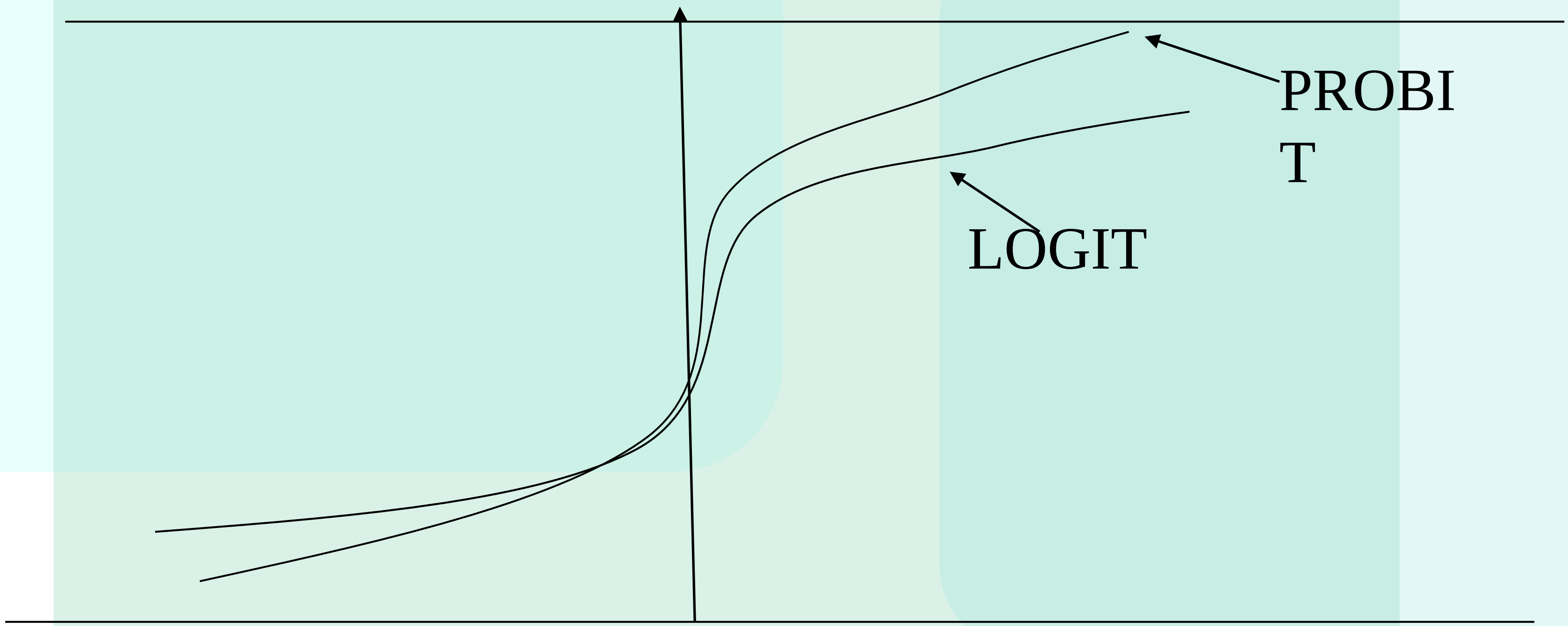
$$P_i = P(Y = 1 | X) = P(I_i^* \leq I_i)$$

Probit Model

$$F(I_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I^*} e^{-z^2/2} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1 + \beta_2 x_i} e^{-z^2/2} dz$$

Probit v.s Logit

- 1



0

Limited Dependent Variable Models and Sample Selection Corrections

- **Limited dependent variables (LDV)**
  - LDV are is substantively restricted
    - Binary vavariables whose range riables, e.g. employed/not employed
    - Nonnegative variables, e.g. wages, prices, interest rates
    - Nonnegative variables with excess zeros, e.g. labor supply
    - Count variables, e.g. the number of arrests in a year
    - Censored variables, e.g. unemployment durations
- **Sample selection models**
  - The sample used to infer population relationships is endogenously selected, e.g. wage offer regression but data only about working women

Limited Dependent Variable Models and Sample Selection Corrections

- **Logit and Probit models for binary response**
- **Disadvantages of the LPM for binary dependent variables**
  - Predictions sometimes lie outside the unit interval
  - Partial effects of explanatory variables are constant
- **Nonlinear models for binary response**
  - Response probability is a nonlinear function of explanat. variables

$$P(y=1|\mathbf{x}) = G(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k) = G(\mathbf{x}\beta)$$

Probability of a "success" given explanatory variables

A cumulative distribution function $0 < G(z) < 1$. The response probability is thus a function of the explanatory variables x.

Shorthand vector notation: the vector of explanatory variables x also contains the constant of the model.

Limited Dependent Variable Models and Sample Selection Corrections

- **Choices for the link function**

Probit: $G(z) = \Phi(z) \equiv \int_{-\infty}^{z} \phi(v)dv$ (normal distribution)

Logit: $G(z) = \Lambda(z) = \exp(z)/[1 + \exp(z)]$ (logistic function)

- **Latent variable formulation of the Logit and Probit models**

$$\boxed{y^* = x\beta + e \quad \text{and} \quad y = 1\,[y^* > 0]}$$

If the latent variable y^* is larger than zero, y takes on the value 1, if it is less or equal zero, y takes on 0 (y^* can thus be interpreted as the propensity to have y = 1)

$$\Rightarrow P(y = 1|\mathbf{x}) = P(y^* > 0|\mathbf{x})$$

$$= P(e > -x\beta) = 1 - G(-x\beta) = G(x\beta)$$

Limited Dependent Variable Models and Sample Selection Corrections

- **Interpretation of coefficients in Logit and Probit models**

Continuous explanatory variables:

$$\boxed{\frac{\partial P(y=1|\mathbf{x})}{\partial x_j}} = g(\mathbf{x}\boldsymbol{\beta})\beta_j \quad \text{where} \quad g(z) \equiv \partial G(z)/\partial z > 0$$

How does the probability for y = 1 change if explanatory variable x_j changes by one unit?

Discrete explanatory variables:

$$G\left[\beta_0 + \beta_1 x_1 + \ldots + \beta_k \boxed{(c_k + 1)}\right] - G\left[\beta_0 + \beta_1 x_1 + \ldots + \beta_k \boxed{c_k}\right]$$

For example, explanatory variable x_k increases by one unit.

- **Partial effects are nonlinear and depend on the level of $\mathbf{x}$**

Limited Dependent Variable Models and Sample Selection Corrections

- **Maximum likelihood estimation of Logit and Probit models**

$$\boxed{f(y_i|\mathbf{x}_i;\boldsymbol{\beta}) = [G(\mathbf{x}_i\boldsymbol{\beta})]^{y_i}\,[1 - G(\mathbf{x}_i\boldsymbol{\beta})]^{1-y_i}}$$

The probability that individual i's outcome is y_i given that his/her characteristics are x_i

$$\log L(\boldsymbol{\beta}) = \log\left(\prod_{i=1}^{n} f(y_i|\mathbf{x}_i;\boldsymbol{\beta})\right) = \sum_{i=1}^{n} \log f(y_i|\mathbf{x}_i;\boldsymbol{\beta})$$

Under random sampling

$$\max\ \log L(\boldsymbol{\beta}) \quad \rightarrow \quad \widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_k$$

= Maximum likelihood estimates

- **Properties of maximum likelihood estimators**
  - Maximum likelihood estimators are consistent, asymptotically normal, and asymptotically efficient if the distributional assumptions hold

Limited Dependent Variable Models and Sample Selection Corrections

- **Hypothesis testing after maximum likelihood estimation**
  - The usual t-tests and confidence intervals can be used
  - There are three alternatives to test multiple hypotheses
    - Lagrange multiplier or score test (not discussed here)
    - Wald test (requires only estimation of unrestricted model)
    - Likelihood ratio test (restricted and unrestricted models needed)

$$LR = 2(\log L_{ur} - \log L_r) \sim \chi_q^2$$

Chi-square distribution with q degrees of freedom

The null hypothesis that the q hypotheses hold is rejected if the growth in maximized likelihood is too large when going from the restricted to the unrestricted model

Limited Dependent Variable Models and Sample Selection Corrections

- **Goodness-of-fit measures for Logit and Probit models**
  - Percent correctly predicted

$$\tilde{y}_i = \begin{cases} 1 & \text{if } G(\mathbf{x}_i\hat{\beta}) \geq .5 \\ 0 & \text{otherwise} \end{cases}$$

Individual i's outcome is predicted as one if the probability for this event is larger than .5, then percentage of correctly predicted y = 1 and y = 0 is counted

  - Pseudo R-squared

$$\tilde{R}^2 = 1 - \log L_{ur} / \log L_0$$

Compare maximized log-likelihood of the model with that of a model that only contains a constant (and no explanatory variables)

  - Correlation based measures

$$Corr(y_i, \tilde{y}_i), \ Corr(y_i, G(\mathbf{x}_i\hat{\beta}))$$

Look at correlation (or squared correlation) between predictions or predicted prob. and true values

Limited Dependent Variable Models and Sample Selection Corrections

- **Reporting partial effects of explanatory variables**
  - The difficulty is that partial effects are not constant but depend on $\mathbf{x}$
  - Partial effects at the average:

$$\widehat{PEA}_j = g(\bar{\boldsymbol{x}}\hat{\boldsymbol{\beta}})\hat{\beta}_j$$

The partial effect of explanatory variable x_j is considered for an "average" individual (this is problematic in the case of explanatory variables such as gender)

  - Average partial effects:

$$\widehat{APE}_j = n^{-1}\sum_{i=1}^{n} g(\mathbf{x}_i\hat{\boldsymbol{\beta}})\hat{\beta}_j$$

The partial effect of explanatory variable x_j is computed for each individual in the sample and then averaged across all sample members (makes more sense)

  - Analogous formulas hold for discrete explanatory variables

Limited Dependent Variable Models and Sample Selection Corrections

- **Example: Married women's labor force participation**

TABLE 17.1 LPM, Logit, and Probit Estimates of Labor Force Participation

Dependent Variable: *inlf*			
Independent Variables	LPM (OLS)	Logit (MLE)	Probit (MLE)
nwifeinc	−.0034 (.0015)	−.021 (.008)	−.012 (.005)
educ	.038 (.007)	.221 (.043)	.131 (.025)
exper	.039 (.006)	.206 (.032)	.123 (.019)
exper^2	−.00060 (.00019)	−.0032 (.0010)	−.0019 (.0006)
age	−.016 (.002)	−.088 (.015)	−.053 (.008)
kidslt6	−.262 (.032)	−1.443 (.204)	−.868 (.119)
kidsge6	.013 (.014)	.060 (.075)	.036 (.043)
constant	.586 (.152)	.425 (.860)	.270 (.509)
Percentage correctly predicted	73.4	73.6	73.4
Log-likelihood value	—	−401.77	−401.30
Pseudo *R*-squared	.264	.220	.221

The coefficients are not comparable across models

Often, Logit estimated coefficients ¼ is 1.6 times Probit estimated because $g_{Logit}(0)/g_{Probit}(0) \approx 1/1.6$.

The biggest difference between the LPM and Logit/Probit is that partial effects are nonconstant in Logit/Probit:

$$\widehat{P}(working|\bar{x}, kidslt6 = 0) = .707$$

$$\widehat{P}(working|\bar{x}, kidslt6 = 1) = .373$$

$$\widehat{P}(working|\bar{x}, kidslt6 = 2) = .117$$

(Larger decrease in probability for the first child)

Limited Dependent Variable Models and Sample Selection Corrections

- **The Tobit model for corner solution responses**
  - In many economic contexts, decision problems are such that either a positive amount or a zero amount is chosen (e.g. demand for alcohol)
  - A linear regression model may be inadequate in such cases as predictions may be negative and effects of explanatory variables are linear
  - The Tobit model also makes use of a latent variable formulation

- **Definition of the Tobit model**

$$y^* = x\beta + u, \quad u|\mathrm{x} \sim \mathrm{Normal}(0, \sigma^2)$$

Conditional on the values of the explanatory variables, the error term is homoskedastic normally distributed

$$y = \max(0, y^*)$$

The final outcome of the dependent variable is positive or zero

Limited Dependent Variable Models and Sample Selection Corrections

- **Maximum likelihood estimation of the Tobit model**

$$f(y_i|\mathbf{x}_i;\boldsymbol{\beta},\sigma) = \begin{cases} \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left[-(y_i - \mathbf{x}_i\boldsymbol{\beta})^2/(2\sigma^2)\right] & \text{if } y_i > 0 \\ 1 - \Phi(\mathbf{x}_i\boldsymbol{\beta}/\sigma) & \text{if } y_i = 0 \end{cases}$$

For positive outcomes, the normal density (applied to $y_i - \mathbf{x}_i\boldsymbol{\beta}$) is used, for zero outcomes the probability is one minus the probability that the latent variable is greater than zero (see Probit).

<u>Maximization of the log-likelihood</u>:

$$\max \ \log L(\boldsymbol{\beta},\sigma) = \sum_{i=1}^{n} \log f(y_i|\mathbf{x}_i;\boldsymbol{\beta},\sigma) \ \rightarrow \ \hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k, \hat{\sigma}$$

As in the Logit/Probit case, the maximization problem is highly nonlinear. It cannot be solved analytically and has to be solved with the help of computer software using e.g. Newton-Raphson methods.

Limited Dependent Variable Models and Sample Selection Corrections

- **Interpretation of the coefficients in the Tobit model**

<u>Conditional mean for all outcomes</u>:

$$E(y|\mathbf{x}) = P(y>0|\mathbf{x})\cdot E(y|y>0,\mathbf{x}) + P(y=0|\mathbf{x})\cdot 0$$

$$= \boxed{\Phi(\boldsymbol{x\beta}/\sigma)\cdot E(y|y>0,\mathbf{x})}$$

The mean for all outcomes is a scaled version of the mean for only the positve outcomes (this is the reason why a regression using only the positive outcomes would yield wrong results)

<u>Conditional mean for positive outcomes</u>:

$$E(y|y>0,\mathbf{x}) = \boldsymbol{x\beta} + \boxed{\sigma\lambda(\boldsymbol{x\beta}/\sigma)}$$

The mean for only the positive outcomes is the usual linear regression but plus an extra term (this is again a reason why an ordinary linear regression would yield wrong results)

$$\lambda(c) = \phi(c)/\Phi(c) > 0$$

This is the socalled <u>inverse Mills ratio</u>

Limited Dependent Variable Models and Sample Selection Corrections

- **Partial effects of interest in the Tobit model**

On the probability for a nonzero outcome:

$$\frac{\partial P(y>0|\mathbf{x})}{\partial x_j} = (\beta_j/\sigma)\phi(\boldsymbol{x\beta}/\sigma)$$

Note that all partial effects depend on the explanatory variables and the error variance

On the mean for positive outcomes:

This adjustment factor can be shown to lie between zero and one

$$\frac{\partial E(y|y>0|\mathbf{x})}{\partial x_j} = \beta_j \overbrace{\{1-\lambda(\boldsymbol{x\beta}/\sigma)\left[\boldsymbol{x\beta}/\sigma+\lambda(\boldsymbol{x\beta}/\sigma)\right]\}}$$

On the mean of all possible outcomes including zero:

$$\frac{\partial E(y|\mathbf{x})}{\partial x_j} = \beta_j\Phi(\boldsymbol{x\beta}/\sigma)$$

Note that this adjustment factor also lies between zero and one

Limited Dependent Variable Models and Sample Selection Corrections

- **Estimation of average partial effects in the Tobit model**

On the probability for a nonzero outcome:

$$\widehat{APE}_{1,j} = n^{-1}\sum_{i=1}^{n}(\hat{\beta}_j/\hat{\sigma})\phi(\mathbf{x}_i\hat{\boldsymbol{\beta}}/\hat{\sigma})$$

Analogous formulas are available for partial effects at the average (PEA) but they have the afore mentioned disadvantages

On the mean for positive outcomes:

$$\widehat{APE}_{2,j} = n^{-1}\sum_{i=1}^{n}\hat{\beta}_j\Big\{1-\lambda(\mathbf{x}_i\hat{\boldsymbol{\beta}}/\hat{\sigma})\Big[\mathbf{x}_i\hat{\boldsymbol{\beta}}/\hat{\sigma}+\lambda(\mathbf{x}_i\hat{\boldsymbol{\beta}}/\hat{\sigma})\Big]\Big\}$$

On the mean of all possible outcomes including zero:

$$\widehat{APE}_{3,j} = n^{-1}\sum_{i=1}^{n}\hat{\beta}_j\Phi(\mathbf{x}_i\hat{\boldsymbol{\beta}}/\hat{\sigma})$$

Limited Dependent Variable Models and Sample Selection Corrections

- **Example: Annual hours worked of married women**

TABLE 17.2 OLS and Tobit Estimation of Annual Hours Worked

Dependent Variable: *hours*

Independent Variables	Linear (OLS)	Tobit (MLE)
nwifeinc	−3.45 (2.24)	−8.81 (4.46)
educ	28.76 (13.04)	80.65 (21.58)
exper	65.67 (10.79)	131.56 (17.28)
$exper^2$	−.700 (.372)	−1.86 (0.54)
age	−30.51 (4.24)	−54.41 (7.42)
kidslt6	−442.09 (57.46)	−894.02 (111.88)
kidsge6	−32.78 (22.80)	−16.22 (38.64)
constant	1,330.48 (274.88)	965.31 (446.44)
Log-likelihood value	—	−3,819.09
R-squared	.266	.274
$\hat{\sigma}$	750.18	1,122.02

Because of the different scaling factors involved, Tobit coefficients are not comparable to OLS coefficients.

To compare Tobit and OLS, one has to compare average partial effects (or partial effects at the average). It turns out that partial effects of Tobit and OLS are different in a number of cases.

Another difference between Tobit and OLS is that, due to the linearity of the model, OLS assumes constant partial effects, whereas partial effects are nonconstant in Tobit.

In the given example, OLS yields negative annual hours for 39 out of 753 women. This is not much but it may be a reason to view the linear model as misspecified.

Limited Dependent Variable Models and Sample Selection Corrections

- **Specification issues in Tobit / Logit / Probit models**
  - A restriction of the Tobit model is that explanat. var. influence positive outcomes and the probability of positive outcomes in the same way
  - This may be unrealistic in many cases, for example, when modeling the relationship between the amount of life insurance and a person's age
  - For such cases, more advanced so-called hurdle models can be used
  - As in Logit/Probit models, heteroskedasticity may be an issue in Tobit

Limited Dependent Variable Models and Sample Selection Corrections

- ML estimates may be wrong if distributional assumptions do not hold
- There are methods to deal with endogeneity in Logit/Probit/Tobit
- Logit/Probit/Tobit models are also available for panel/time series data

Limited Dependent Variable Models and Sample Selection Corrections

- **The Poisson regression model for count data**

$$P(y=h) = \exp[-\mu]\,[\mu]^h / h!,\ \ h = 0, 1, 2, \ldots$$

Probability that y takes on the integer value h

Probability function of the Poisson distribution, where $\mu = E(y) > 0$

Model the mean of the dependent variable as a function of explanatory variables:

$$\mu(\mathbf{x}) = \exp(\boldsymbol{x\beta}) = \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k) > 0$$

The Poisson regression model models a count variable as a function of explanatory variables:

$$P(y=h|\mathbf{x}) = \exp[-\exp(\boldsymbol{x\beta})]\,[\exp(\boldsymbol{x\beta})]^h / h!,\ \ h = 0, 1, 2, \ldots$$

Limited Dependent Variable Models and Sample Selection Corrections

- **Interpretation of the coefficients of the Poisson regression**

$$\frac{\partial \mu(\mathbf{x})}{\partial x_j} = \exp(\boldsymbol{x}\boldsymbol{\beta})\beta_j = \mu(\mathbf{x})\beta_j \Rightarrow \boxed{\beta_j = \frac{\frac{\partial \mu(\mathbf{x})}{\mu(\mathbf{x})}}{\partial x_j}}$$

By what percentage does the mean outcome change if x_j is increased by one?

- **Maximum likelihood estimation of the Poisson regression model**

$$\max \log L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \log P(y = y_i|\mathbf{x}_i) = \sum_{i=1}^{n} y_i \mathbf{x}_i \boldsymbol{\beta} - \exp(\mathbf{x}_i \boldsymbol{\beta})$$

$$E(y|\mathbf{x}) = Var(y|\mathbf{x})$$

- A limitation of the model is that it assumes
- But ML estimators in the Poisson regression model are consistent and asymptotically normal even if the Poisson distribution does not hold

Limited Dependent Variable Models and Sample Selection Corrections

- **Poisson regression for number of arrests**

TABLE 17.3 Determinants of Number of Arrests for Young Men

Dependent Variable: *narr86*

Independent Variables	Linear (OLS)	Exponential (Poisson QMLE)
pcnv	−.132 (.040)	−.402 (.085)
avgsen	−.011 (.012)	−.024 (.020)
tottime	.012 (.009)	.024 (.015)
ptime86	−.041 (.009)	−.099 (.021)
qemp86	−.051 (.014)	−.038 (.029)
inc86	−.0015 (.0003)	−.0081 (.0010)
black	.327 (.045)	.661 (.074)
hispan	.194 (.040)	.500 (.074)
born60	−.022 (.033)	−.051 (.064)
constant	.577 (.038)	−.600 (.067)
Log-likelihood value	—	−2,248.76
R-squared	.073	.077
$\hat{\sigma}$	.829	1.232

The expected number of arrests was 2.4 percentage points lower if the average sentence length was 1 month higher.

If the assumption of a Poisson distribution does not hold, ML is still consistent and asymptotically normal. This is called Quasi-Maximum Likelihood estimation (QML).

If the distributional assumptions do not hold and QML is used, standard errors are wrong. One then has to compute robust standard errors (this has not been done here).

Because $\hat{\sigma} = 1.232$ there is evidence for overdispersion:

$$Var(y|\mathbf{x}) = \sigma^2 E(y|\mathbf{x}) \approx 1.518\ E(y|\mathbf{x})$$

This is evidence that the Poisson distribution does not hold and robust standard errors have to be computed. Alternatively one can inflate standard errors by $\hat{\sigma}$ the factor.

Limited Dependent Variable Models and Sample Selection Corrections

- **The censored regression model**
  - In many cases, the dependent variable is censored in the sense that values are only reported up to a certain level (e.g. top coded wealth)
  - Censored normal regression model:

True outcome (unobserved)

$$y_i = \mathbf{x}_i\beta + u_i, \quad u_i|\mathbf{x}_i, c_i \sim \text{Normal}(0, \sigma^2)$$

Observed outcome

$$w_i = \min(y_i, c_i)$$

If the true outcome exceeds the censoring threshold, only the threshold is reported

  - Regressing y_i on x_i would yield correct results but y_i is unobserved
  - Regressing w_i on x_i will yield incorrect results (even if only the un-censored observations are used in this regression)

Limited Dependent Variable Models and Sample Selection Corrections

- **Maximum likelihood estimation of the censored regression model**

$$P(w_i = c_i|\mathbf{x}_i) = P(y_i \geq c_i|\mathbf{x}_i)$$

$$= P(u_i \geq c_i - \mathbf{x}_i\beta) = 1 - \Phi\left[(c_i - \mathbf{x}_i\beta)/\sigma\right]$$

If the censoring threshold does not bind, the density of the outcome is normal

Probability/density function of observed outcome conditional on explanatory variables:

$$f(w_i|\mathbf{x}_i;\beta,\sigma) = \begin{cases} \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left[-(w_i - \mathbf{x}_i\beta)^2/(2\sigma^2)\right] & \text{if } w_i < c_i \\ 1 - \Phi((c_i - \mathbf{x}_i\beta)/\sigma) & \text{if } w_i = c_i \end{cases}$$

Maximization of log-likelihood:

$$\max \log L(\beta,\sigma) = \sum_{i=1}^{n} \log f(w_i|\mathbf{x}_i, c_i) \quad \rightarrow \quad \hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k, \hat{\sigma}$$

Limited Dependent Variable Models and Sample Selection Corrections

- **Censored regression estimation of criminal recidivism**

TABLE 17.4 Censored Regression Estimation of Criminal Recidivism

Dependent Variable: log(*durat*)	
Independent Variables	Coefficient (Standard Error)
workprg	−.063 (.120)
priors	−.137 (.021)
tserved	−.019 (.003)
felon	.444 (.145)
alcohol	−.635 (.144)
drugs	−.298 (.133)
black	−.543 (.117)
married	.341 (.140)
educ	.023 (.025)
age	.0039 (.0006)
constant	4.099 (.348)
Log-likelihood value $\hat{\sigma}$	−1,597.06 1.810

The variable durat measures the time in months until a prison inmate is arrested after being released from prison. Of 1,445 inmates, 893 had not been arrested during the time they were followed. Their time out of prison is censored (because its end, if there was one, was not observed).

For example, if the time in prison was one month longer, this reduced the expected duration until the next arrest by about 1.9%.

<u>In the censored regression model, the coefficients can be directly interpreted.</u> This is contrary to the Tobit model, where coefficients cannot be directly interpreted. The censored regression model and the Tobit model have a similar structure, but in the Tobit model, the outcome is of a nonlinear nature whereas in the censored regression model, the outcome is linear but incompletely observed.

Limited Dependent Variable Models and Sample Selection Corrections

- **Truncated regression models**
  - In a truncated regression model the outcome and the explanatory variables are only observed if the outcome is less or equal some value c_i
  - In this case, the sample is not a random sample from the population (because some units will never be a part of the sample)
  - Truncated normal regression model:

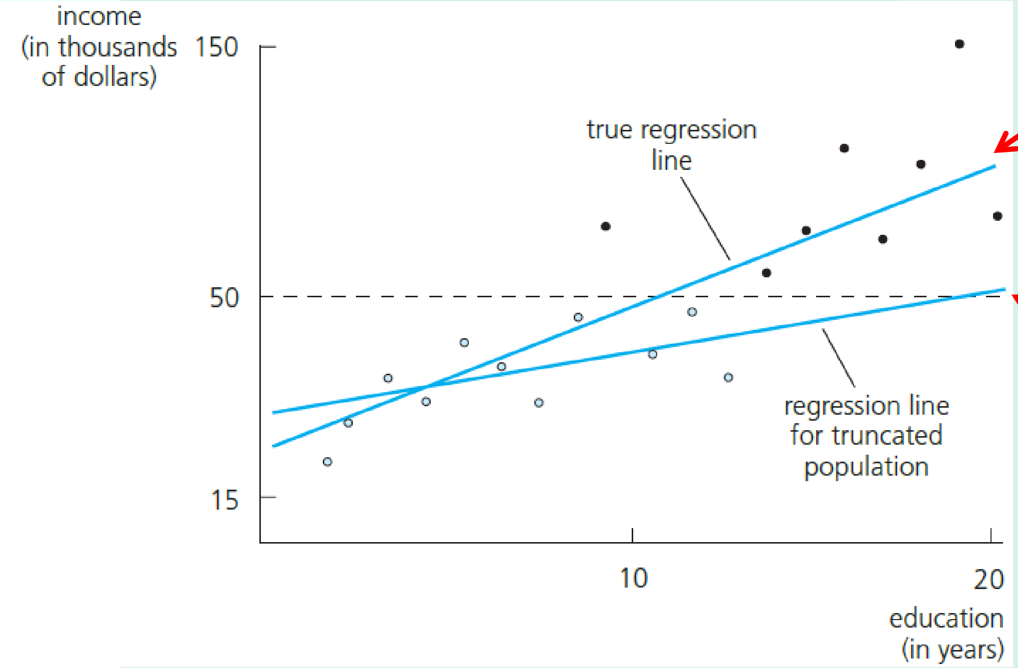
$$y_i = \mathbf{x}_i\beta + u_i,\ \ u_i|\mathbf{x}_i \sim \text{Normal}(0,\sigma^2)$$

$(\mathbf{x}_i, y_i)$ only observed if $y_i \leq c_i$

  - Applying OLS would not yield correct results because MLR.2 is violated

Limited Dependent Variable Models and Sample Selection Corrections

- **Example: A regression based on a truncated sample**



One is interested in the relationship between income and education in the whole population but sampling was done so that only individuals with income below $50,000 were sampled.

Direct application of OLS would lead to an underestimation of the slope coefficient because high incomes are omitted.

Limited Dependent Variable Models and Sample Selection Corrections

- **Maximum likelihood estimation of the truncated regression model**

Density of an observed outcome conditional on explanatory variables and the threshold c_i:

$$g(y_i|\mathbf{x}_i, c_i) = \frac{f(y_i|\mathbf{x}_i\boldsymbol{\beta}, \sigma^2)}{P(y_i \leq c_i|\mathbf{x}_i)} = \frac{f(y_i|\mathbf{x}_i\boldsymbol{\beta}, \sigma^2)}{F(c_i|\mathbf{x}_i\boldsymbol{\beta}, \sigma^2)}$$

Density and distribution function of a normal distribution with mean $\mathbf{x}_i\boldsymbol{\beta}$ and variance σ^2

Likelihood maximization:

$$\max \log L(\boldsymbol{\beta}, \sigma) = \sum_{i=1}^{n} \log g(y_i|\mathbf{x}_i, c_i) \quad \rightarrow \quad \widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_k, \widehat{\sigma}$$

- **As in the censored regression model, nonnormality or heteroskedasticity in the truncated regression model lead to inconsistency**

Limited Dependent Variable Models and Sample Selection Corrections

- **Sample selection corrections**
  - The question is under which assumptions a sample with nonrandom sample selection can be used to infer relationships in the population

- **When is OLS on the selected sample consistent?**

$$y = \mathbf{x}\beta + u, \;\; E(u|\mathbf{x}) = 0$$ Population model

s_i Sample selection indicator, s_i = 1 if observation is part of the sample, s_i = 0 otherwise

$$s_i y_i = s_i \mathbf{x}_i \beta + s_i u_i$$ Regression based on selected sample

Condition for consistency of OLS: $Corr(sx_j, su) = E(sx_j u) = 0$

Limited Dependent Variable Models and Sample Selection Corrections

- **Three cases in which OLS on the selected sample is consistent**
  - Selection is independent of explanatory variables and the error term
  - Selection is completely determined by explanatory variables
  - Selection depends on the explanatory variables and other factors that are uncorrelated with the error term
- **Similar results apply to IV/2SLS estimation**
  - Instead of for explanatory variables, the conditions have to hold for the full list of exogenous variables that are used in the model
- **Sample selection and nonlinear models estimated by ML**
  - Consistency if sample selection is only determined by explanat. var.

Limited Dependent Variable Models and Sample Selection Corrections

- **Incidental truncation (Heckman model)**
- **Example: Wage offer function using sample of working women**
  - One is interested in the wage of a woman with certain characteristics would be offered on the labor market if she decided to work
  - Unfortunately, one only observes the wages of women who actually work, i.e. who have accepted the wage offered to them
  - The sample is truncated because women who do not work (but who would be offered a wage if they asked for it) are never in the sample
  - Truncation of this kind is called incidental truncation because it depends on another variable (here: labor force participation)

Limited Dependent Variable Models and Sample Selection Corrections

- **Definition of Heckman model**

$$y = \mathbf{x}\beta + u$$

Main equation (e.g. wage equation)

$$s = 1\,[\mathbf{z}\gamma + v \geq 0]$$

Selection equation (e.g. whether working)

$$(u, v)|\mathbf{x}, \mathbf{z} \sim \text{Normal}(0, \rho)$$

The error terms of both equations are jointly normally distributed (independent of the explanatory variables) with correlation coefficient ρ

- **Selection bias in OLS**

$$E(y|\mathbf{x}, \mathbf{z}, s = 1) = \mathbf{x}\beta + \boxed{\rho\lambda(\mathbf{z}\gamma)}$$

For example, if the correlation of unobserved wage determinants and unobserved determinants of the work decision is positive, the women observed working will have higher wages than expected from their characteristics (= positive selection)

Running the regression on the truncated sample suffers from omitted variable bias

Limited Dependent Variable Models and Sample Selection Corrections

- **Estimation of Heckman model**

1) Estimation of correction term:

$$P(s=1|\mathbf{z}) = \Phi(\mathbf{z}\gamma)$$

Estimate Probit for work decision using all observations (working and nonworking women)

$$\hat{\lambda} = \phi(z\hat{\gamma})/\Phi(z\hat{\gamma})$$

Calculate inverse Mills ratio using Probit coefficients

2) Include estimated correction term in regression:

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + \rho\hat{\lambda} + error$$

If this coefficient is different from zero, there is nonignorable sample selection bias

There have to be explanatory variables in the selection equation that are not in the main equation (exclusion restrictions), otherwise there is multicollinearity because the inverse Mills ratio is almost linear in z

Limited Dependent Variable Models and Sample Selection Corrections

- **Example: Wage offer equation for married women**

TABLE 17.5 Wage Offer Equation for Married Women

Dependent Variable: log(*wage*)		
Independent Variables	OLS	Heckit
educ	.108 (.014)	.109 (.016)
exper	.042 (.012)	.044 (.016)
$exper^2$	−.00081 (.00039)	−.00086 (.00044)
constant	−.522 (.199)	−.578 (.307)
$\hat{\lambda}$	—	.032 (.134)
Sample size *R*-squared	428 .157	428 .157

The standard errors of the two-step Heckman method are actually wrong and have to be corrected (not done here). One can also use a maximum likelihood procedure.

There is no significant sample selection. This is the reason why OLS and Heckman estimates ("Heckit") are so similar.