Applied Econometrics

Lectures 13 & 14:

Nonlinear Models Beyond Binary Choice:

Multinomial Response Models, Corner Solution Models &

Censored Regressions

Måns Söderbom[*]

6 & 9 October 2009

[*] University of Gothenburg. Email: mans.soderbom@economics.gu.se. Web: http://www.soderbom.net

1. Introduction

These notes refer to two lectures in which we consider the following econometric models:

- Multinomial response models (e.g. whether an individual is unemployed, wage-employed or self-employed)
- Ordered response models (e.g. modelling the rating of the corporate payment default risk, which varies from, say, A (best) to D (worst))
- Corner solution models and censored regression models (e.g. modelling household health expenditure: the dependent variable is non-negative, continuous above zero and has a lot of observations at zero)

These models are designed for situations in which the dependent variable is not strictly continuous and not binary.

References:

Wooldridge (2002): 15.9 (Multinomial response); 15.10 (Ordered response); 16.1-5; 16.6.3-4; 16.7; 17.3 (Corner solutions and censoring).

2. Ordered Response Models

What's the meaning of **ordered response**? Consider credit rating on a scale from zero to six, for instance, and suppose this is the variable that we want to model (i.e. this is the dependent variable). Clearly, this is a variable that has ordinal meaning: six is better than five, which is better than four etc.

The standard way of modelling ordered response variables is by means of **ordered probit** or **ordered logit**. These two models are very similar. I will discuss the ordered probit, but everything below carries over to the logit if we replace the normal CDF $\Phi(.)$ by the logistic CDF $\Lambda(.)$.

- Can you think of reasons why OLS may not be suitable for modelling an ordered response variable?
- Could binary choice models (LPM, probit, logit) potentially be used?

2.1. Ordered Probit

Let y be an ordered response taking on the values $\{0, 1, 2, ..., J\}$. We derive the ordered probit from a **latent variable model** (cf. probit binary choice)

$$\begin{aligned} y^* &= \beta_1 x_1 + ... + \beta_k x_k + e \\ &= \boldsymbol{x\beta} + e, \end{aligned} \tag{2.1}$$

where e is a normally distributed variable with the variance normalized to one. Notice that this model does **not** contain a constant. Next define J **cut-off points** (or **threshold parameters**) as follows:

$$\alpha_1 < \alpha_2 < ...\alpha_J.$$

We do not observe the latent variable, but we do observe choices according to the following:

$$\begin{aligned} y &= 0 \text{ if } y^* \leq \alpha_1 \\ y &= 1 \text{ if } \alpha_1 < y^* \leq \alpha_2 \\ y &= 2 \text{ if } \alpha_2 < y^* \leq \alpha_3 \\ & (...) \\ y &= J \text{ if } \alpha_J < y^*. \end{aligned}$$

Suppose y can take three values: 0, 1 or 2 (Wooldridge, 2002, pp.504-508 provides an exposition of the general case with J unspecified). We then have

$$\begin{aligned} y &= 0 \text{ if } \boldsymbol{x\beta} + e \leq \alpha_1 \\ y &= 1 \text{ if } \alpha_1 < \boldsymbol{x\beta} + e \leq \alpha_2 \\ y &= 2 \text{ if } \alpha_2 < \boldsymbol{x\beta} + e. \end{aligned}$$

We can now define the probabilities of observing $y = 0, 1, 2$. For the smallest and the largest value, the resulting expressions are very similar to what we have seen for the binary probit:

$$\begin{aligned}
\Pr\left(y=0|x\right) &= \Pr\left(\boldsymbol{x\beta}+e\leq\alpha_1\right) \\
&= \Pr\left(e\leq\alpha_1-\boldsymbol{x\beta}\right) \\
&= \Phi\left(\alpha_1-\boldsymbol{x\beta}\right), \\
&= 1-\Phi\left(\boldsymbol{x\beta}-\alpha_1\right) \\
\Pr\left(y=2|x\right) &= \Pr\left(\boldsymbol{x\beta}+e>\alpha_2\right) \\
&= \Pr\left(e>\alpha_2-\boldsymbol{x\beta}\right) \\
&= 1-\Phi\left(\alpha_2-\boldsymbol{x\beta}\right) \\
&= \Phi\left(\boldsymbol{x\beta}-\alpha_2\right).
\end{aligned}$$

For the intermediate category, we get:

$$\begin{aligned}
\Pr\left(y=1|x\right) &= \Pr\left(\alpha_1<\boldsymbol{x\beta}+e\leq\alpha_2\right) \\
&= \Pr\left(e>\alpha_1-\boldsymbol{x\beta}, e\leq\alpha_2-\boldsymbol{x\beta}\right) \\
&= \left[1-\Phi\left(\alpha_1-\boldsymbol{x\beta}\right)\right]-\Phi\left(\boldsymbol{x\beta}-\alpha_2\right), \\
&= 1-\left(1-\Phi\left(\boldsymbol{x\beta}-\alpha_1\right)\right)-\Phi\left(\boldsymbol{x\beta}-\alpha_2\right), \\
&= \Phi\left(\boldsymbol{x\beta}-\alpha_1\right)-\Phi\left(\boldsymbol{x\beta}-\alpha_2\right),
\end{aligned}$$

or equivalently

$$\Pr\left(y=1|x\right)=\Phi\left(\alpha_2-\boldsymbol{x\beta}\right)-\Phi\left(\alpha_1-\boldsymbol{x\beta}\right)$$

(remember: $\Phi\left(a\right)=1-\Phi\left(-a\right)$, because the normal distribution is symmetric - keep this in mind when studying ordered probits or you might get lost in the algebra). In the general case where there are several intermediate categories, all the associated probabilities will be of this form; see Wooldridge (2002), p.506.

Notice that the probabilities sum to one.

2.2. Interpretation

When discussing binary choice models we paid a lot of attention to marginal effects - i.e. the partial effects of a small change in explanatory variable x_j on the probability that we have a positive outcome.

For ordered models, we can clearly compute marginal effects on the predicted probabilities along the same principles. It is not obvious (to me, anyway) that this the most useful way of interpreting the results is, however. Let's have a look the marginal effects and then discuss.

2.2.1. Partial effects on predicted probabilities

When discussing marginal effects for binary choice models, we focussed on the effects on the probability that y (the binary dependent variable) is equal to one. We ignored discussing effects on the probability that y is equal to zero, as these will always be equal to minus one times the partial effect on the probability that y is equal to one.

Since we now have more than two outcomes, interpretation of partial effects on probabilities becomes somewhat more awkward. Sticking to the example in which we have three possible outcomes, we obtain:

$$\frac{\partial \Pr(y=2|x)}{\partial x_k} = \phi\left(\boldsymbol{x\beta} - \alpha_2\right)\beta_k,$$

for the highest category (note: analogous to the expression for binary probit).[1] Moreover,

$$\frac{\partial \Pr(y=1|x)}{\partial x_k} = \left[\phi\left(\boldsymbol{x\beta} - \alpha_1\right) - \phi\left(\boldsymbol{x\beta} - \alpha_2\right)\right]\beta_k,$$

for the intermediate category, and

[1] Remember that $\phi(a) = \phi(-a)$ - i.e. I could just as well have written

$$\frac{\partial \Pr(y=2|x)}{\partial x_k} = \phi\left(\alpha_2 - \beta x\right)\beta_k,$$

for instance - cf. Wooldridge's (2002) exposition on pp. 505-6).

$$\frac{\partial \Pr(y=0|x)}{\partial x_k} = -\phi(\boldsymbol{x\beta} - \alpha_1)\,\beta_k,$$

for the lowest category, assuming that x_k is a continuous variable enter the index model linearly (if x_k is discrete - typically binary - you just compute the discrete change in the predicted probabilities associated with changing x_k by one unit, for example from 0 to 1). We observe:

- The partial effect of x_k on the predicted probability of the highest outcome has the same sign as β_k.
- The partial effect of x_k on the predicted probability of the lowest outcome has the opposite sign to β_k
- The sign of the partial effect of x_k on predicted probabilities of intermediate outcomes cannot, in general, be inferred from the sign of β_k. This is because there are two offsetting effects - suppose $\beta_k > 0$, then the intermediate category becomes more likely if you increase x_k because the the probability of the lowest category falls, but it also becomes less likely because the the probability of the highest category increases (illustrate this in a graph). Typically, partial effects for intermediate probabilities are quantitatively small and often statistically insignificant. Don't let this confuse you!

Discussion - how best interpret results from ordered probit (or logit)?

- Clearly one option here is to look at the estimated β-parameters, emphasizing the underlying **latent variable equation** with which we started. Note that we don't identify the standard deviation of e separately. Note also that consistent estimation of the β-parameters requires the model to be correctly specified - e.g. homoskedasticity and normality need to hold, if we are using ordered probit. Such assumptions are testable using, for example, the methods introduced for binary choice models. You don't often see this done in applied work however.
- Another option might be to look at the effect on the **expected value** of the ordered response

variable, e.g.

$$\frac{\partial E\left(y|x,\beta\right)}{\partial x_k}=\frac{\partial \Pr\left(y=0|x\right)}{\partial x_k}\times 0+\frac{\partial \Pr\left(y=1|x\right)}{\partial x_k}\times 1+\frac{\partial \Pr\left(y=2|x\right)}{\partial x_k}\times 2,$$

in our example with three possible outcomes. This may make a lot of sense if y is a numerical variable - basically, if you are prepared to compute mean values of y and interpret them. For example, suppose you've done a survey measuring consumer satisfaction where 1="very unhappy", 2="somewhat unhappy", 3="neither happy nor unhappy", 4="somewhat happy", and 5="very happy", then most people would be prepared to look a the sample mean even though strictly the underlying variable is qualitative, thinking that 3.5 (for example) means something (consumers are on average a little bit happy?). In such a case you could look at partial effects on the conditional mean.

- Alternatively, you might want investigate the effect on the probability of observing categories $j, j+1, ..., J$. In my consumer satisfaction example, it would be straightforward to compute the partial effect on the probability that a consumer is "somewhat happy" or "very happy", for example.
- Thus, it all boils down to presentation and interpretation here, and exactly what your quantity of interest is depends on the context. We can use the Stata command 'mfx compute' to obtain estimates of the partial effects on the predicted probabilities, but for more elaborate partial effects you may have to do some coding tailored to the context.

EXAMPLE: Incidence of corruption in Kenyan firms. Section 1 in the appendix.

3. Multinomial response: Multinomial logit

Suppose now the dependent variable is such that more than two outcomes are possible, where the outcomes cannot be ordered in any natural way. For example, suppose we are modelling occupational status based on household data, where the possible outcomes are self-employed (SE), wage-employed (WE) or unemployed (UE). Alternatively, suppose we are modelling the transportation mode for commuting to work: bus, train, car,...

Binary probit and logit models are ill suited for modelling data of this kind. Of course, in principle we could bunch two or more categories and so construct a binary outcome variable from the raw data (e.g. if we don't care if employed individuals are self-employed or wage-employees, we may decide to construct a binary variable indicating whether someone is unemployed or employed). But in doing so, we throw away potentially interesting information. And OLS is obviously not a good model in this context.

However, the logit model for binary choice can be **extended** to model more than two outcomes. Suppose there are J possible outcomes in the data. The dependent variable y can then take J values, e.g. 0,1,...,J-1. So if we are modelling, say, occupational status, and this is either SE, WE or UE, we have $J = 3$. There is no natural ordering of these outcomes, and so what number goes with what category is arbitrary (but, as we shall see, it matters for the interpretation of the results). Suppose we decide on the following:

$$\begin{aligned} y &= 0 \text{ if individual is UE,} \\ y &= 1 \text{ if individual is WE,} \\ y &= 2 \text{ if individual is SE.} \end{aligned}$$

We write the conditional probability that an individual belongs to category $j = 0, 1, 2$ as

$$\Pr\left(y_i = j | \boldsymbol{x}_i\right),$$

where $\boldsymbol{x}_i$ is a vector of explanatory variables. Reasonable restrictions on these probabilities are:

- that each of them is bounded in the (0,1) interval,
- that they sum to unity (one).

One way of imposing these restrictions is to write the probabilities in logit form:

$$\Pr(y_i = 1|\boldsymbol{x}_i) = \frac{\exp(\boldsymbol{x}_i\boldsymbol{\beta}_1)}{1+\exp(\boldsymbol{x}_i\boldsymbol{\beta}_1)+\exp(\boldsymbol{x}_i\boldsymbol{\beta}_2)},$$

$$\Pr(y_i = 2|\boldsymbol{x}_i) = \frac{\exp(\boldsymbol{x}_i\boldsymbol{\beta}_2)}{1+\exp(\boldsymbol{x}_i\boldsymbol{\beta}_1)+\exp(\boldsymbol{x}_i\boldsymbol{\beta}_2)},$$

$$\begin{aligned}\Pr(y_i = 0|\boldsymbol{x}_i) &= 1-\Pr(y_i = 1|\boldsymbol{x}_i)-\Pr(y_i = 2|\boldsymbol{x}_i)\\ &= \frac{1}{1+\exp(\boldsymbol{x}_i\boldsymbol{\beta}_1)+\exp(\boldsymbol{x}_i\boldsymbol{\beta}_2)}.\end{aligned}$$

The main difference compared to what we have seen so far, is that there are now **two** parameter vectors, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ (in the general case with J possible responses, there are $J-1$ parameter vectors). This makes interpretation of the coefficients more difficult than for binary choice models.

- The easiest case to think about is where β_{1k} and β_{2k} have the same sign. If β_{1k} and β_{2k} are positive (negative) then it is clear that an increase in the variable x_k makes it less (more) likely that the individual belongs to category 0.

- But what about the effects on $\Pr(y_i = 1|\boldsymbol{x}_i)$ and $\Pr(y_i = 2|\boldsymbol{x}_i)$? This is much trickier than what we are used to. We know that, for sure, the sum of $\Pr(y_i = 1|\boldsymbol{x}_i)$ and $\Pr(y_i = 2|\boldsymbol{x}_i)$ will increase, but how this total increase is allocated between these two probabilities is not obvious. To find out, we need to look at the marginal effects. We have

$$\begin{aligned}\frac{\partial \Pr(y_i = 1|\boldsymbol{x}_i)}{\partial x_{ik}} &= \beta_{1k}\exp(\boldsymbol{x}_i\boldsymbol{\beta}_1)\left[1+\exp(\boldsymbol{x}_i\boldsymbol{\beta}_1)+\exp(\boldsymbol{x}_i\boldsymbol{\beta}_2)\right]^{-1}\\ &\quad -\exp(\boldsymbol{x}_i\boldsymbol{\beta}_1)\left[1+\exp(\boldsymbol{x}_i\boldsymbol{\beta}_1)+\exp(\boldsymbol{x}_i\boldsymbol{\beta}_2)\right]^{-2}\\ &\quad \times\left(\beta_{1k}\exp(\boldsymbol{x}_i\boldsymbol{\beta}_1)+\beta_{2k}\exp(\boldsymbol{x}_i\boldsymbol{\beta}_2)\right),\end{aligned}$$

which can be written as

$$\begin{aligned}\frac{\partial \Pr\left(y_i=1|\boldsymbol{x}_i\right)}{\partial x_{ik}} &= \beta_{1k}\Pr\left(y_i=1|\boldsymbol{x}_i\right) \\ &-\Pr\left(y_i=1|\boldsymbol{x}_i\right)\left[1+\exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_1\right)+\exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_2\right)\right]^{-1} \\ &\times\left(\beta_{1k}\exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_1\right)+\beta_{2k}\exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_2\right)\right),\end{aligned}$$

or

$$\frac{\partial \Pr\left(y_i=1|\boldsymbol{x}_i\right)}{\partial x_{ik}}=\Pr\left(y_i=1|\boldsymbol{x}_i\right)\left[\beta_{1k}-\frac{\beta_{1k}\exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_1\right)+\beta_{2k}\exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_2\right)}{1+\exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_1\right)+\exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_2\right)}\right]. \tag{3.1}$$

Similarly, for $j=2$:

$$\frac{\partial \Pr\left(y_i=2|\boldsymbol{x}_i\right)}{\partial x_{ik}}=\Pr\left(y_i=2|\boldsymbol{x}_i\right)\left[\beta_{2k}-\frac{\beta_{1k}\exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_1\right)+\beta_{2k}\exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_2\right)}{1+\exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_1\right)+\exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_2\right)}\right],$$

while for the base category $j=0$:

$$\frac{\partial \Pr\left(y_i=0|\boldsymbol{x}_i\right)}{\partial x_{ik}}=\Pr\left(y_i=0|\boldsymbol{x}_i\right)\left[-\frac{\beta_{1k}\exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_1\right)+\beta_{2k}\exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_2\right)}{1+\exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_1\right)+\exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_2\right)}\right].$$

Of course it's virtually impossible to remember, or indeed interpret, these expressions. The point is that whether the probability that y falls into, say, category 1 rises or falls as a result of varying x_{ik}, depends not only on the parameter estimate β_{1k}, but also on β_{2k}. As you can see from (3.1), the marginal effect $\frac{\partial \Pr(y_i=1|\boldsymbol{x}_i)}{\partial x_{ik}}$ may in fact be negative even if β_{1k} is positive, and vice versa. Why might that happen?

- EXAMPLE: Appendix, Section 2. Occupational outcomes amongst Kenyan manufacturing workers.

3.0.2. Independence of irrelevant alternatives (IIA)

The multinomial logit is very convenient for modelling an unordered discrete variable that can take on more than two values. One important limitation of the model is that the ratio of any two probabilities j

and m depends **only** on the parameter vectors β_j and β_m, and the explanatory variables x_i:

$$\begin{aligned}\frac{\Pr(y_i = 1|\boldsymbol{x}_i)}{\Pr(y_i = 2|\boldsymbol{x}_i)} &= \frac{\exp(\boldsymbol{x}_i\boldsymbol{\beta}_1)}{\exp(\boldsymbol{x}_i\boldsymbol{\beta}_2)} \\ &= \exp(\boldsymbol{x}_i(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)).\end{aligned}$$

It follows that the inclusion or exclusion of **other categories** must be irrelevant to the ratio of the two probabilities that $y = 1$ and $y = 2$. This is potentially restrictive, in a behavioral sense.

Example: Individuals can commute to work by three transportation means: blue bus, red bus, or train. Individuals choose one of these alternatives, and the econometrician estimates a multinomial logit modelling this decision, and obtains an estimate of

$$\frac{\Pr(y_i = \text{red bus}|\boldsymbol{x}_i)}{\Pr(y_i = \text{train}|\boldsymbol{x}_i)}.$$

Suppose the bus company were to remove blue bus from the set of options, so that individuals can choose only between red bus and train. If the econometrician were to estimate the multinomial logit on data generated under this regime, do you think the above probability ratio would be the same as before?

If not, this suggests the multinomial logit modelling the choice between blue bus, red bus and train is mis-specified: the presence of a blue bus alternative is not irrelevant for the above probability ratio, and thus for individuals' decisions more generally.

Some authors (e.g. Greene; Stata manuals) claim we can **test** the IIA assumption for the multinomial logit by means of a Hausman test. The basic idea is as follows:

1. Estimate the full model. For example, with red bus, train and blue bus being the possible outcomes, and with red bus defined as the benchmark category. Retain the coefficient estimates.

2. Omit one category and re-estimate the model - e.g. exclude blue bus, and model the binary decision to go by train as distinct from red bus.

3. Compare the coefficients from (1) and (2) above using the usual Hausman formula. Under the null

that IIA holds, the coefficients should not be significantly different from each other.

Actually this procedure does not make sense.

- First, you don't really have data generated in the alternative regime (with blue bus not being an option) and so how can you hope to shed light on the behavioral effect of removing blue bus from the set of options?

- Second, obviously **sample means** of ratios such as

$$\frac{\Pr(y_i = \text{red bus})}{\Pr(y_i = \text{train})} = \frac{N_{\text{red bus}}/N}{N_{\text{train}}/N}$$

  don't depend on blue bus outcomes. So if you estimate a multinomial logit with only a constant included in the specification, the estimated constant in the specification train specification (with red bus as the reference outcome) will not change if you omit blue bus outcomes when estimating (i.e. step (2) above). Conceptually, a similar issue will hold if you have explanatory variables in the model, at least if you have a flexible functional form in your $\boldsymbol{x}_i\boldsymbol{\beta}$ indices (e.g. mutually exclusive dummy variables)

- Third, from what I have seen the Hausman test for the IIA does not work well in practice (not very surprising).

- Finally, note the Wooldridge discusses the IIA in the context of **conditional** logit models, i.e. models where choices are made based on observable attributes of each alternative (e.g. ticket prices for blue bus, red bus and train may vary). For such a model the test for IIA makes good sense. What I have said above applies specifically for the multinomial logit.

Note that there are lots of other econometric models that can be used to model multinomial response models - notably multinomial probit, conditional logit, nested logit etc. These will not be discussed here.

EXAMPLE: Hausman test for IIA based on multinomial logit gives you nonsense - appendix, Section 3.

4. Corner Solution Models

We now consider econometric issues that arise when the dependent variable is bounded but continuous within the bounds. We focus first on corner solution models, and then turn to the censored regression model (duration data is often censored) and truncated regression.

In general, a **corner solution response variable** is bounded such that

$$lo \leq y_i \leq hi,$$

where lo denotes the lower bound (limit) and hi the higher bound, and where these bounds are the result of real economic constraints.

- By far the most common case is $lo = 0$ and $hi = \infty$, i.e. there is a lower limit at zero and no upper limit. The dependent variable takes the value zero for a nontrivial fraction of the population, and is roughly continuously distributed over positive values. You will often find this in micro data, e.g. household expenditure on education (Kingdon, 2005), health, alcohol,... or investment in capital goods among small entrepreneurial firms (Bigsten et al., 2005).

- You can thus think of this type of variable as a **hybrid** between a continuous variable (for which the linear model is appropriate) and a binary variable (for which one would typically use a binary choice model). Indeed, as we shall see, the econometric model designed to model corner solution variables looks like a hybrid between OLS and the probit model. In what follows we focus on the case where $lo = 0$, $hi = \infty$, however generalizing beyond this case is reasonably straightforward.

Let y be a variable that is equal to zero for some non-zero proportion of the population, and that is continuous and positive if it is not equal to zero. As usual, we want to model y as a function of a set of variables $x_1, x_2, ..., x_k$ - or in matrix notation:

$$\boldsymbol{x} = \begin{bmatrix} 1 & x_1 & x_2 & ... & x_k \end{bmatrix}.$$

4.1. OLS

We have seen how for binary choice models OLS can be a useful starting point (yielding the linear probability model), even though the dependent variable is not continuous. We now have a variable which is 'closer' to being a continuous variable - it's discrete in the sense that it is either in the corner (equal to zero) or not (in which case it's continuous).

OLS is a useful starting point for modelling corner solution variables:

$$y = \boldsymbol{x\beta} + u.$$

We've seen that there are a number of reasons why we may not prefer to estimate binary choice models using OLS. For similar reasons OLS may not be an ideal estimator for corner response models:

- Based on OLS estimates we can get **negative predictions**, which doesn't make sense since the dependent variable is non-negative (if we are modelling household expenditure on education, for instance, negative predicted values do not make sense).

- Conceptually, the idea that a corner solution variable is **linearly** related to a continuous independent variable for all possible values is a bit suspect. It seems more likely that for observations close to the corner (close to zero), changes in some continuous explanatory variable (say x_1) has a smaller effect on the outcome than for observations far away from the corner. So if we are interested in understanding how y depends on x_1 among low values of y, linearity is not attractive.

- A third (and less serious) problem is that the residual u is likely to be heteroskedastic - but we can deal with this by simply correcting the standard errors.

- A fourth and related problem is that, because the distribution of y has a 'spike' at zero, the residual cannot be normally distributed. This means that OLS point estimates are unbiased, but inference in small samples cannot be based on the usual suite of normality-based distributions such as the t test.

So you see all of this is very similar to the problems identified with the linear probability model.

4.2. Tobit

To fix these problems we follow a similar path as for binary choice models. We start, however, from the latent variable model, written as

$$y^* = \boldsymbol{x}\boldsymbol{\beta} + u, \tag{4.1}$$

where the residual u is assumed **normally distributed** with a **constant variance** σ_u^2, and uncorrelated with $\boldsymbol{x}$. Exogeneity can be relaxed using techniques similar to those adopted for probit models with endogenous regressors, see Section 16.2.2 in Wooldridge (2002). As usual, the latent variable y^* is unobserved - we observe

$$y = \left\{ \begin{array}{c} y^* \text{ if } y^* > 0 \\ 0 \text{ if } y^* \leq 0 \end{array} \right\}, \tag{4.2}$$

which can be written equivalently as

$$y = \max\left(y^*, 0\right).$$

Two things should be noted here.

- First, y^* satisfies the classical linear model assumptions, so had y^* been observed the obvious choice of estimator would have been OLS.

- Second, it is often helpful to think of y as a variable that is bounded below for **economic** reasons, and y^* as a variable that reflects the 'desired' value if there were no constraints. Actual household expenditure on health is one example - this is bounded below at zero. In such a case y^* could be interpreted as desired expenditure, in which case $y^* < 0$ would reflect a desire to sell off ones personal (or family's) health. This may not be as far-fetched as it sounds - if you're very healthy and very poor, for instance, perhaps you wouldn't mind feeling a little less healthy if you got paid for it (getting paid here, of course, would be the same as having negative health expenditure).

We said above that a corner solution variable is a kind of hybrid: both discrete and continuous. The discrete part is due to the piling up of observations at zero. The probability that y is equal to zero can be written

$$\begin{aligned}
\Pr\left(y=0|x\right) &= \Pr\left(y^{*}\leq 0\right), \\
&= \Pr\left(\boldsymbol{x\beta}+u\leq 0\right), \\
&= \Pr\left(u\leq -\boldsymbol{x\beta}\right) \\
&= \Phi\left(\frac{-\boldsymbol{x\beta}}{\sigma_u}\right) \text{ (integrate; normal distribution)} \\
\Pr\left(y=0|x\right) &= 1-\Phi\left(\frac{\boldsymbol{x\beta}}{\sigma_u}\right) \text{ (by symmetry),}
\end{aligned}$$

exactly like the probit model. In contrast, if $y>0$ then it is continuous:

$$y=\boldsymbol{x\beta}+u.$$

It follows that the conditional density of y is equal to

$$f\left(y|\boldsymbol{x};\boldsymbol{\beta},\sigma_u\right)=\left[1-\Phi\left(\boldsymbol{x}_i\boldsymbol{\beta}/\sigma_u\right)\right]^{1_{[y(i)=0]}}\left[\phi\left(\frac{y_i-\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma_u}\right)\right]^{1_{[y(i)>0]}},$$

where $1_{[a]}$ is a dummy variable equal to one if a is true. Thus the contribution of observation i to the sample log likelihood is

$$\ln L_i=1_{[y(i)=0]}\ln\left[1-\Phi\left(\boldsymbol{x}_i\boldsymbol{\beta}/\sigma_u\right)\right]+1_{[y(i)>0]}\ln\left[\phi\left(\frac{y_i-\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma_u}\right)\right],$$

and the sample log likelihood is

$$\ln L\left(\boldsymbol{\beta},\sigma_u\right)=\sum_{i=1}^{N}\ln L_i.$$

Estimation is done by means of maximum likelihood.

4.2.1. Interpreting the tobit model

Suppose the model can be written according to the equations (4.1)-(4.2), and suppose we have obtained estimates of the parameter vector

$$\beta = \left[\begin{array}{ccccc} \beta_0 & \beta_1 & \beta_2 & ... & \beta_k \end{array} \right].$$

How do we interpret these parameters?

We see straight away from the latent variable model that β_j is interpretable as the partial (marginal) effects of x_j on the latent variable y^*, i.e.

$$\frac{\partial E\left(y^*|\boldsymbol{x}\right)}{\partial x_j} = \beta_j,$$

if x_j is a continuous variable, and

$$E\left(y^*|x_j = 1\right) - E\left(y^*|x_j = 0\right) = \beta_j$$

if x_j is a dummy variable (of course if x_j enters the model nonlinearly these expressions need to be modified accordingly). I have omitted i-subscripts for simplicity. If that's what we want to know, then we are home: all we need is an estimate of the relevant parameter β_j.

Typically, however, we are interested in the partial effect of x_j on the expected **actual outcome** y, rather than on the latent variable. Think about the health example above. We are probably primarily interested in the partial effects of x_j (perhaps household size) on expected actual - rather than desired - health expenditure, e.g. $\partial E\left(y|\boldsymbol{x}\right)/\partial x_j$ if x_j is continuous. In fact there are two different potentially interesting marginal effects, namely

$$\frac{\partial E\left(y|\boldsymbol{x}\right)}{\partial x_j}, \qquad \text{(Unconditional on y)}$$

and

$$\frac{\partial E\left(y|\boldsymbol{x},y>0\right)}{\partial x_j}. \qquad \text{(Conditional on y>0)}$$

We need to be clear on which of these we are interested in. Now let's see what these marginal effects look like.

The marginal effects on expected y, conditional on y positive. We want to derive

$$\frac{\partial E\left(y|\boldsymbol{x},y>0\right)}{\partial x_j}.$$

Recall that the model can be written

$$\begin{aligned} y &= \max\left(y^*,0\right), \\ y &= \max\left(\boldsymbol{x\beta}+u,0\right) \end{aligned}$$

(see (4.1)-(4.2)). We begin by writing down $E\left(y|x,y>0\right)$:

$$\begin{aligned} E\left(y|y>0,\boldsymbol{x}\right) &= E\left(\boldsymbol{x\beta}+u|y>0,\boldsymbol{x}\right), \\ E\left(y|y>0,\boldsymbol{x}\right) &= \boldsymbol{x\beta}+E\left(u|y>0,\boldsymbol{x}\right), \\ E\left(y|y>0,\boldsymbol{x}\right) &= \boldsymbol{x\beta}+E\left(u|u>-\boldsymbol{x\beta}\right) \end{aligned}$$

Because of the truncation (y is always positive, or, equivalently, u is always larger than $-\boldsymbol{x\beta}$), dealing with the second term is not as easy as it may seem. We begin by taking on board the following result for normally distributed variables:

- **A useful result.** If z follows a normal distribution with mean zero, and variance equal to one (i.e. a standard normal distribution), then

$$E\left(z|z>c\right)=\frac{\phi\left(c\right)}{1-\Phi\left(c\right)}, \tag{4.3}$$

where c is a constant (i.e. the lower bound here), ϕ denotes the standard normal probability density, and Φ is the standard normal cumulative density.

The residual u is not, in general, standard normal because the variance is not necessarily equal to one, but by judiciously dividing and multiplying through with its standard deviation σ_u we can transform u to become standard normal:

$$E\left(y|y>0,x\right)=\boldsymbol{x\beta}+\sigma_u E\left(u/\sigma_u|u/\sigma_u>-\boldsymbol{x\beta}/\sigma_u\right).$$

That is, (u/σ_u) is now standard normal, and so we can apply the above 'useful result', i.e. eq (4.3), and write:

$$E\left(u|u>-\boldsymbol{x\beta}\right)=\sigma_u\frac{\phi\left(-\boldsymbol{x\beta}/\sigma_u\right)}{1-\Phi\left(-\boldsymbol{x\beta}/\sigma_u\right)},$$

and thus

$$E\left(y|y>0,\boldsymbol{x}\right)=\boldsymbol{x\beta}+\sigma_u\frac{\phi\left(-\boldsymbol{x\beta}/\sigma_u\right)}{1-\Phi\left(-\boldsymbol{x\beta}/\sigma_u\right)}.$$

With slightly cleaner notation,

$$E\left(y|y>0,\boldsymbol{x}\right)=\boldsymbol{x\beta}+\sigma_u\frac{\phi\left(\boldsymbol{x\beta}/\sigma_u\right)}{\Phi\left(\boldsymbol{x\beta}/\sigma_u\right)},$$

which is often written as

$$E\left(y|y>0,\boldsymbol{x}\right)=\boldsymbol{x\beta}+\sigma_u\lambda\left(\boldsymbol{x\beta}/\sigma_u\right), \tag{4.4}$$

where the function λ is defined as

$$\lambda\left(z\right)=\frac{\phi\left(z\right)}{\Phi\left(z\right)}.$$

in general, and known as the **inverse Mills ratio** function.

- Have a look at the inverse Mills ratio function in Section 4 in the appendix, Figure 1.

Equation (4.4) shows that the expected value of y, given that y is not zero, is equal to $\boldsymbol{x\beta}$ **plus** a term $\sigma_u \lambda\left(\boldsymbol{x\beta}/\sigma_u\right)$ which is strictly positive (how do we know that?).

We can now obtain the marginal effect:

$$\begin{aligned}
\frac{\partial E\left(y|y>0,\boldsymbol{x}\right)}{\partial x_j} &= \beta_j + \sigma_u \frac{\partial \lambda\left(\boldsymbol{x\beta}/\sigma_u\right)}{\partial x_j}, \\
&= \beta_j + \sigma_u\left(\beta_j/\sigma_u\right)\lambda', \\
&= \beta_j\left(1+\lambda'\right),
\end{aligned}$$

where λ' denotes the partial derivative of λ with respect to $\left(\boldsymbol{x\beta}/\sigma_u\right)$ (note: I am assuming here that x_j is continuous and not functionally related to any other variable - i.e. it enters the model linearly - this means I can use calculus, and that I don't have to worry about higher-order terms). It is tedious but fairly easy to show that

$$\lambda'\left(z\right) = -\lambda\left(z\right)\left[z+\lambda\left(z\right)\right]$$

in general, hence

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

This shows that the partial effect of x_j on $E\left(y|y>0,\boldsymbol{x}\right)$ is not determined just by β_j. In fact, it depends on **all parameters $\boldsymbol{\beta}$** in the model as well as on the values of **all explanatory variables $\boldsymbol{x}$**, and the standard deviation of the residual. The term in $\{\cdot\}$ is often referred to as the **adjustment factor**, and it can be shown that this is always larger than zero and smaller than one (why is this useful to know?).

It should be clear that, just as in the case for probits and logits, we need to evaluate the marginal effects at specific values of the explanatory variables. This should come as no surprise, since one of the reasons we may prefer tobit to OLS is that we have reasons to believe the marginal effects may differ according to how close to the corner (zero) a given observation is (see above). In Stata we can use the *mfx compute* command to compute marginal effects without too much effort. How this is done will be clearer in a moment, but first I want to go over the second type of marginal effect that I might be interested in.

The marginal effects on expected y, unconditional on the value of y Recall:

$$
\begin{aligned}
y &= \max\left(y^{*}, 0\right), \\
y &= \max\left(\boldsymbol{x\beta} + u, 0\right).
\end{aligned}
$$

I now need to derive

$$
\frac{\partial E\left(y|\boldsymbol{x}\right)}{\partial x_j}.
$$

We write $E\left(y|\boldsymbol{x}\right)$ as follows:

$$
\begin{aligned}
E\left(y|x\right) &= \Phi\left(-\boldsymbol{x\beta}/\sigma_u\right) \cdot E\left(y|y=0,\boldsymbol{x}\right) + \Phi\left(\boldsymbol{x\beta}/\sigma_u\right) \cdot E\left(y|y>0,\boldsymbol{x}\right), \\
&= \Phi\left(-\boldsymbol{x\beta}/\sigma_u\right) \cdot 0 + \Phi\left(\boldsymbol{x\beta}/\sigma_u\right) \cdot E\left(y|y>0,\boldsymbol{x}\right), \\
&= \Phi\left(\boldsymbol{x\beta}/\sigma_u\right) \cdot E\left(y|y>0,\boldsymbol{x}\right),
\end{aligned}
$$

i.e. the probability that y is positive times the expected value of y given that y is indeed positive. Using the product rule for differentiation,

$$
\frac{\partial E\left(y|x\right)}{\partial x_j} = \Phi\left(\boldsymbol{x\beta}/\sigma_u\right) \cdot \frac{\partial E\left(y|y>0,\boldsymbol{x}\right)}{\partial x_j} + \phi\left(\boldsymbol{x\beta}/\sigma_u\right)\frac{\beta_j}{\sigma_u} \cdot E\left(y|y>0,\boldsymbol{x}\right),
$$

and we know from the previous sub-section that

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

and

$$
E\left(y|y>0,\boldsymbol{x}\right) = \boldsymbol{x\beta} + \sigma_u \lambda\left(\boldsymbol{x\beta}/\sigma_u\right).
$$

Hence

$$\begin{aligned}\frac{\partial E\left(y|\boldsymbol{x}\right)}{\partial x_j} &= \Phi\left(\boldsymbol{x\beta}/\sigma_u\right)\cdot\beta_j\left\{1-\lambda\left(\boldsymbol{x\beta}/\sigma_u\right)\left[\boldsymbol{x\beta}/\sigma_u+\lambda\left(\boldsymbol{x\beta}/\sigma_u\right)\right]\right\} \\ &\quad +\phi\left(\boldsymbol{x\beta}/\sigma_u\right)\frac{\beta_j}{\sigma_u}\cdot\left[\boldsymbol{x\beta}+\sigma_u\lambda\left(\boldsymbol{x\beta}/\sigma_u\right)\right],\end{aligned}$$

which looks complicated but the good news is that several of the terms cancel out, so that:

$$\frac{\partial E\left(y|\boldsymbol{x}\right)}{\partial x_j}=\beta_j\Phi\left(\boldsymbol{x\beta}/\sigma_u\right)$$

(try to prove this). This has a straightforward interpretation: the marginal effect of x_j on the expected value of y, conditional on the vector $\boldsymbol{x}$, is simply the parameter β_j times the probability that y is larger than zero. Of course, this probability is smaller than one, so it follows immediately that the marginal effect is strictly smaller than the parameter β_j.

Now consider the example in section 4 in the appendix, on investment in plant and machinery among Ghanaian manufacturing firms.

4.2.2. Specification issues

The choice between $y=0$ vs. $y>0$, and the amount of y given $y>0$, are determined by a single mechanism. One assumption underlying the tobit model when applied to corner solution outcomes is that if some variable x_j impacts, say, positively on the **expected value** of y, given $y>0$, then the **probability** that y is equal to one is also positively related to x_j. In other words, x_j is a **single mechanism** determining both these outcomes. Recall:

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

where the adjustment factor $\{\cdot\}$ is larger than zero and smaller than one, and

$$\frac{\partial \Pr\left(y>0|\boldsymbol{x}\right)}{\partial x_j} = \phi\left(\boldsymbol{x\beta}/\sigma_u\right)\beta_j.$$

This can sometimes be too restrictive: it may be that the probability that $y > 0$ depends negatively on x_j while the expected value of y, given $y > 0$, depends positively on x_j (e.g. life insurance coverage as a function of age: people might be more likely to have life insurance as they get older, but the value of the policies might decrease with age). This possibility is not allowed for in the tobit model.

One way of informally investigating whether this assumption appears accepted by the data or not, is to compare the tobit results to what you get if you estimate a probit model where the (binary) dependent variable is equal to one if $y > 0$ and zero if $y = 0$. The probit model would be specified as

$$\Pr\left(y>0|\boldsymbol{x}\right) = \Phi\left(\boldsymbol{x\gamma}/\sigma_u\right).$$

Recall that neither γ_j nor σ_u is identified separately from the probit model; all we can hope to identify is $\psi_j = \gamma_j/\sigma_u$. Here, however, we have an estimate of σ_u from the tobit model, so under the null hypothesis that the tobit is correct we would have

$$\psi_j = \beta_j/\sigma_u.$$

Of course in practice this will never hold exactly due to sampling error, and so the issue is whether or not these two terms are 'close' or not. For instance, if the probit coefficient ψ_j is positive and significant while the tobit coefficient β_j is negative and significant, this would signal problems. In the section below entitled 'Hurdle models' we consider a generalized estimator that does not suffer from this problem.

- See example based on the Ghana data in handout, page 4.

Heteroskedasticity and non-normality. If the error term u is heteroskedastic (i.e. the variance of u is not constant) and/or non-normal, the tobit model may yield badly biased parameter estimates. If, as is the premise in this section, the dependent variable is a corner response variable, we have already discussed how we're mainly interested in the marginal effects rather than the parameters $\boldsymbol{\beta}$. Heteroskedasticity and non-normality imply that the expressions for $E(y|y>0,x)$ and $E(y|x)$ derived above no longer are correct, and so the above expressions for the marginal effects $\frac{\partial E(y|x)}{\partial x_j}$ and $\frac{\partial E(y|x,y>0)}{\partial x_j}$ are also wrong. In general, this can be interpreted as a problem with the functional form of the model (e.g. the normal CDF and PDF are wrong).

- There are several ways of testing for heteroskedasticity and non-normality in the tobit model. The 'parametric and formal' test discussed for probit models is easy to implement for tobit as well, and will shed some light on the validity the functional form. See Table 5, page 4, in the handout for such a test based on the Ghana data. A more sophisticated approach might be to adopt **conditional moments tests**, which examine whether the relevant sample moments are supported by the data.[2] Normality of u, for example, implies:

  $$E\left[(y^* - \boldsymbol{x\beta})^3\right] = 0$$

  and

  $$E\left[(y^* - \boldsymbol{x\beta})^4 - 3\sigma^4\right] = 0$$

  The problem here is that we do not observe y^*, and so the tests are based on **generalized residuals** - please refer to Tauchen (1985) for details. I have coded Stata programs that implement conditional moment tests for heteroskedasticity and normality, for the probit and tobit model - these can be obtained from my web page (under Resources)..

- If we conclude there is a functional form problem, then one obvious thing to try is to generalize the functional form of the econometric model, perhaps by adding higher-order terms (e.g. squared terms or interaction terms) to the set of explanatory variables.

- Some people prefer alternative estimators, often the **censored least absolute deviations** (CLAD)

[2] See: Newey, W. K. (1985). "Maximum likelihood specification testing and conditional moment tests," *Econometrica* 53, pp. 1047-1073; and Tauchen, G. (1985). "Diagnostic testing and evaluation of maximum likelihood models," *Journal of Econometrics* 30, pp. 415-443.

estimator or non-parametric estimators, to the tobit model, on the grounds that these are more robust to the problems just discussed. We now turn to the CLAD estimator.

4.3. The CLAD estimator

Consider again the latent variable model but with zero **median** of u given $\boldsymbol{x}$:

$$y^* = \boldsymbol{x\beta} + u,$$

$$Med\,(u|\boldsymbol{x}) = 0.$$

No further distributional assumptions are needed. We now bring into play a useful result from probability theory which says that, if $g\,(y)$ is a nondecreasing function, then $Med\,(g\,(y)) = g\,(Med\,(y))$. In our case,

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

hence

$$\begin{aligned} Med\,(y|\boldsymbol{x}) &= Med\,(\max\,(0, y^*)\,|\boldsymbol{x}) \\ Med\,(y|\boldsymbol{x}) &= \max\,(0, Med\,(y^*|\boldsymbol{x})) \\ Med\,(y|\boldsymbol{x}) &= \max\,(0, \boldsymbol{x\beta})\,. \end{aligned}$$

Thus, $Med\,(u|\boldsymbol{x}) = 0$ implies that the median of y conditional on $\boldsymbol{x}$ is equal to zero if $\boldsymbol{x\beta} \leq 0$, and equal to $\boldsymbol{x\beta}$ if $\boldsymbol{x\beta} > 0$. In other words, the (unknown) parameter vector $\boldsymbol{\beta}$ dictates how the median of y conditional on $\boldsymbol{x}$ varies with $\boldsymbol{x}$**.** We can estimate $\boldsymbol{\beta}$ by minimizing the following criterion function:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} |y_i - \max\,(0, \boldsymbol{x}_i\boldsymbol{\beta})|$$

That is, we are minimizing the sum of absolute deviations. Why is that an appropriate criterion function?

In my experience, the CLAD estimator **may** be useful, though note that it can be computationally difficult to implement and often give quite imprecise results (high standard errors). Wooldridge provides a good discussion of this model in Section 16.6.4.

4.4. Hurdle models

If we conclude that the single mechanism assumption is inappropriate, what do we do? One way of proceeding is to estimate a **hurdle model**, where the 'hurdle' is whether or not to choose positive y.[3] Unlike the tobit model, hurdle models separate the initial decision of $y > 0$ versus $y = 0$ from the decision of how much y given $y > 0$.

A simple and useful hurdle model for a corner solution variable is

$$\Pr\left(y > 0 | \boldsymbol{x}\boldsymbol{\gamma}\right) = \Phi\left(\boldsymbol{x}\boldsymbol{\gamma}\right)$$

$$\log y |\left(\boldsymbol{x}, y > 0\right) \sim Normal\left(\boldsymbol{x}\boldsymbol{\beta}, \sigma^2\right).$$

The first equation says that the probability of a positive outcome (overcoming the 'hurdle') is modelled as a probit, while the second equation states that, conditional on a positive outcome, and conditional on the vector of explanatory variables $\boldsymbol{x}$, the dependent variable follows a log normal distribution. It follows that

$$E\left(y|\boldsymbol{x}, y > 0\right) = \exp\left(\boldsymbol{x}\boldsymbol{\beta}+\frac{1}{2}\sigma^2\right),$$

and

$$E\left(y|\boldsymbol{x}\right) = \Phi\left(\boldsymbol{x}\boldsymbol{\gamma}\right)\exp\left(\boldsymbol{x}\boldsymbol{\beta}+\frac{1}{2}\sigma^2\right).$$

Estimation of this model is straightforward:

1. First, estimate $\boldsymbol{\gamma}$ using probit, in which the dependent variable is one if $y > 0$ and zero if $y = 1$. This gives us an estimate of the probability that $y > 0$, conditional on x.

[3] Wooldridge (2002), Chapter 16.7.

2. Second, estimate $\boldsymbol{\beta}$ using a linear regression (e.g. OLS) in which $\ln y$ is the dependent variable, but where observations for which $y = 0$ are excluded. This gives us an estimate of the expected value of y conditional on $y > 0$ and x.

This is quite flexible in that we allow for different mechanisms determining the predicted probability that y is zero vs. non-zero on the one hand, and the expected amount of y, given $y > 0$, on the other. Marginal effects are straightforward to compute:

$$\frac{\partial E\left(y|\boldsymbol{x}, y > 0\right)}{\partial x_k} = \beta_k \exp\left(\boldsymbol{x\beta}+\frac{1}{2}\sigma^2\right),$$

$$\frac{\partial E\left(y|\boldsymbol{x}\right)}{\partial x_k} = \left[\gamma_k \phi\left(\boldsymbol{x\gamma}\right) + \beta_k \Phi\left(\boldsymbol{x\gamma}\right)\right] \exp\left(\boldsymbol{x\beta}+\frac{1}{2}\sigma^2\right),$$

and standard errors may be calculated by means of the delta method or bootstrapping.

Of course, hurdle models can be used even if log normality does not hold. Cragg (1971) considered the case where y, conditional on $\boldsymbol{x}$,$y > 0$, follows a **truncated normal distribution**. In this case the density of y, conditional on $\boldsymbol{x}$ and $y > 0$ is equal to

$$f\left(y|\boldsymbol{x}; y > 0\right) = \frac{\phi\left(\left(y - \boldsymbol{x\beta}\right)/\sigma\right)/\sigma}{\Phi\left(\boldsymbol{x\beta}/\sigma\right)},$$

and so the density of y conditional on $\boldsymbol{x}$ and the unknown parameters of the model becomes

$$f\left(y|\boldsymbol{x}; \boldsymbol{\beta}, \boldsymbol{\gamma}\right) = \left[1 - \Phi\left(\boldsymbol{x\gamma}\right)\right]^{1[y=0]} \left[\Phi\left(\boldsymbol{x\gamma}\right) \frac{\phi\left(\left(y - \boldsymbol{x\beta}\right)/\sigma\right)/\sigma}{\Phi\left(\boldsymbol{x\beta}/\sigma\right)}\right]^{1[y>0]}.$$

Notice that this nests the tobit density of y:

$$f\left(y|\boldsymbol{x}; \boldsymbol{\beta}\right) = \left[1 - \Phi\left(\boldsymbol{x\beta}/\sigma\right)\right]^{1[y=0]} \left[\phi\left(\left(y - \boldsymbol{x\beta}\right)/\sigma\right)/\sigma\right]^{1[y>0]}.$$

Hence, if in Cragg's hurdle model $\boldsymbol{\gamma} = \boldsymbol{\beta}/\sigma$ we have the tobit model. In this hurdle model the contribution of observation i to the sample log likelihood is

$$\ln L_i = 1_{[y_i=0]} \left[1 - \Phi\left(\boldsymbol{x}_i\boldsymbol{\gamma}\right)\right] + 1_{[y_i>0]} \left\{\ln \Phi\left(\boldsymbol{x}_i\boldsymbol{\gamma}\right) - \ln \Phi\left(\boldsymbol{x}_i\boldsymbol{\beta}/\sigma\right) + \ln\left[\phi\left(\left(y_i - \boldsymbol{x}_i\boldsymbol{\beta}\right)/\sigma\right)/\sigma\right]\right\}.$$

This is the log likelihood of the probit plus the log likelihood of the (conditional) truncated y, an insight we can use to form a test of the null hypothesis that the single mechanism assumption underlying the tobit model is supported by the data.

5. Censored and Truncated Models

We have just covered in some detail the tobit model as applied to corner solution models. Recall that a corner solution is an actual economic outcome, e.g. zero expenditure on health by a household in a given period. In this section we discuss briefly two close cousins of the corner solution model, namely the censored regression model and the truncated regression model. The good news is that the econometric techniques used for censored and truncated dependent variables are very similar to what we have already studied.

5.1. Censored regression models

In contrast to corner solutions, censoring is essentially a **data problem**. Censoring occurs, for example, if whenever y exceeds some upper threshold c the actual value of y gets **recorded** as equal to c, rather than the true value. Of course, censoring may also occur at the lower end of the dependent variable. **Top coding** in income surveys is the most common example of censoring, however. Such surveys are sometimes designed so that that people with incomes higher than some upper threshold, say $\$500,000$, are allowed to respond "more than $\$500,000$". In contrast, for people with incomes lower than $\$500,000$ the actual income gets recorded. If we want to run a regression explaining income based on such data, we clearly need to deal with the top coding. A reasonable way of writing down the model might be

$$y^* = \boldsymbol{x}\boldsymbol{\beta} + u,$$

$$y = \min\left(y^{*}, c\right),$$

where y^{*} is **actual** income (which is not fully observed due to the censoring), u is a normally distributed and homoskedastic residual, and y is measured income, which in this example is bounded above at $c = \$500,000$ due to the censoring produced by the design of the survey.

You now see that the censored regression is very similar to the corner solution model. In fact, if $c = 0$ and this is a lower bound, the econometric model for corner solution models and censored regressions coincide: in both cases we would have the tobit model. If the threshold c is not zero and/or represents an upper rather than a lower bound on what is observed, then we still use tobit but with a simple (and uninteresting) adjustment of the log likelihood.

The only substantive difference between censored regressions models and corner solution models lies in the **interpretation of the results**. Suppose we have two models:

- Model 1: the dependent variable is a corner solution variable, with the corner at zero
- Model 2: the dependent variable is censored below at zero.

We could use exactly the same econometric estimator for both models, i.e. the tobit model. In the corner solution model we are probably mainly interested in how the expected value of the observed dependent variable varies with the explanatory variable(s). This means we should look at $E\left(y|\boldsymbol{x}, y > 0\right)$ or $E\left(y|\boldsymbol{x}\right)$, and we have seen in the previous section how to obtain the relevant marginal effects. However, for the censored regression model we are mostly interested in learning how the expected value of the **unobserved and censored** variable y^{*} varies with the explanatory variable(s), i.e. $E\left(y^{*}|\boldsymbol{x}\right)$:

$$E\left(y^{*}|\boldsymbol{x}\right) = \boldsymbol{x\beta},$$

and so the partial effect of x_j is simply β_j.

5.1.1. Duration Data

One field in which censored regression models are very common is in the econometric analysis of **duration data.** Duration is the time that elapses between the 'beginning' and the 'end' of some specified state. The most common example is unemployment duration, where the 'beginning' is the day the individual becomes unemployed and the 'end' is when the same individual gets a new job. Other examples are the duration of wars, duration of marriages, time between first and second child, the lifetimes of firms, the length of stay in graduate school, time to adoption of new technologies, length of financial crises etc etc.

Data on durations are often censored, either to the right (common) or to the left (not so common) or both (even less common). Right censoring means that we don't know from the data when a certain duration ended; left censoring means that we don't know when it began. I will not cover duration data as part of this course, but you can find an old lecture introducing duration data models on my web page.

5.2. Truncated regression models

A truncated regression model is similar to a censored regression model, but there is one important difference:

- If the dependent variable is truncated we do not observe **any** information about a certain segment in the population.
- In other words, we do not have a representative (random) sample from the population. This can happen if a survey targets a sub-group of the population. For instance when surveying firms in developing countries, the World Bank often excludes firms with less than 10 employees. Clearly if we are modelling employment based on such data we need to recognize the fact that firms with less than 10 employees are not covered in our dataset.
- Alternatively, it could be that we target poor individuals, and so exclude everyone with an income higher than some upper threshold c.

- The standard truncated regression model is written

$$y = \boldsymbol{x\beta} + u,$$

where the residual u is assumed normally distributed, homoskedastic and uncorrelated with $\boldsymbol{x}$ (the latter assumption can be relaxed if we have instruments). Suppose that all observations for which $y_i > c$ are excluded from the sample. Our objective is to estimate the parameter $\boldsymbol{\beta}$.

- See example in appendix, Section 5.

It is clear from the example in the appendix that ignoring the truncation leads to substantial downward bias in the estimate of $\boldsymbol{\beta}$. Fortunately, we can correct this bias fairly easily, by using the normality assumption in combination with the information about the threshold. The density of y, conditional on $\boldsymbol{x}$ and y observed, takes a familiar form:

$$f\left(y|\boldsymbol{x};\boldsymbol{\beta},\boldsymbol{\gamma}\right) = \left[\frac{\phi\left(\left(y-\boldsymbol{x\beta}\right)/\sigma\right)/\sigma}{\Phi\left(\boldsymbol{x\beta}/\sigma\right)}\right],$$

and the individual log likelihood contribution is

$$\ln L_i = \ln\left[\phi\left(\left(y_i-\boldsymbol{x}_i\boldsymbol{\beta}\right)/\sigma\right)/\sigma\right] - \ln\Phi\left(\boldsymbol{x}_i\boldsymbol{\beta}/\sigma\right)$$

The conditional expected value of y is also of a familiar form:

$$E\left(y|y>0,\boldsymbol{x}\right) = \boldsymbol{x\beta} + \sigma_u\lambda\left(\boldsymbol{x\beta}/\sigma_u\right)$$

In Stata we can implement this model using the **truncreg** command (see appendix).

PhD Programme: Applied Econometrics
Department of Economics, University of Gothenburg
Appendix: Lectures 13 & 14
Måns Söderbom

1. Ordered probit: Incidence of corruption among Kenyan manufacturing firms

In the following example we consider a model of corruption in the Kenyan manufacturing sector.[1] Our dataset consists of 155 firms observed in year 2000.

Our basic latent model of corruption is

$$corrupt_i^* = \alpha_1 \ln K_i + \alpha_2 \left(\frac{profit}{K} \right)_i + s_i + town_i + e_i \,,$$

where
corrupt = incidence of corruption in the process of getting connected to public services
K = Value of the firm's capital stock
profit = Total profit
s = sector effect (food, wood, textile; metal is the omitted base category)
town = location effect (Nairobi, Mombasa, Nakuru; Eldoret – which is the most remote town – is the omitted base category)
u = a residual, assumed homoskedastic and normally distributed with variance normalized to one.

Incidence of corruption is not directly observed. Instead we have subjective data, collected through interviews with the firm's management, on the prevalence of corruption. Specifically, each firm was asked the following question:

"Do firms like yours typically need to make extra, unofficial payments to get connected to public services (e.g. electricity, telephone etc)?"

Answers were coded using the following scale:

N/A	Always	Usually	Frequently	Sometimes	Seldom	Never
0	1	2	3	4	5	6

Observations for which the answer is N/A or missing have been deleted from the data. Notice that this variable, denoted *obribe*, is ordered so that high values indicate relatively low levels of corruption.

Given the data available, it makes sense to estimate the model using either ordered probit or ordered logit.

[1] These data was collected by a team from the CSAE in 2000 – for details on the survey and the data, see Söderbom, Måns "Constraints and Opportunities in Kenyan Manufacturing: Report on the Kenyan Manufacturing Enterprise Survey 2000," 2001, CSAE Report REP/2001-03. Oxford: Centre for the Study of African Economies, Department of Economics, University of Oxford. Available at http://www.economics.ox.ac.uk/CSAEadmin/reports.

Summary statistics for these variables are as follows:

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      obribe |       155    3.154839    1.852138          1          6
          lk |       155    15.67499    3.197098   7.258711   22.38821
       profk |       155   -.3647645    2.449862  -15.73723    11.3445
        wood |       155          .2    .4012966          0          1
     textile |       155    .2903226    .4553826          0          1
-------------+--------------------------------------------------------
       metal |       155    .2516129    .4353465          0          1
     nairobi |       155    .5096774    .5015268          0          1
     mombasa |       155    .2645161     .442505          0          1
      nakuru |       155    .1032258    .3052398          0          1
```

Table 1. Ordered probit results

```
.  oprobit obribe1 lk profk sec2-sec4 nairobi mombasa nakuru

Iteration 0:   log likelihood = -257.79967
Iteration 1:   log likelihood = -248.35111
Iteration 2:   log likelihood = -248.34599
Iteration 3:   log likelihood = -248.34599

Ordered probit estimates                          Number of obs   =        155
                                                  LR chi2(8)      =      18.91
                                                  Prob > chi2     =     0.0154
Log likelihood = -248.34599                       Pseudo R2       =     0.0367

------------------------------------------------------------------------------
     obribe1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          lk |  -.0809392   .0307831    -2.63   0.009     -.141273   -.0206054
       profk |  -.0569773   .0377651    -1.51   0.131    -.1309955    .0170409
        wood |   -.543739   .2698032    -2.02   0.044    -1.072543   -.0149345
     textile |   .1068028   .2405553     0.44   0.657    -.3646768    .5782825
       metal |  -.3959804    .251102    -1.58   0.115    -.8881313    .0961706
     nairobi |   .0740607   .2836262     0.26   0.794    -.4818364    .6299578
     mombasa |  -.1443718   .3005436    -0.48   0.631    -.7334265    .4446829
      nakuru |  -.0242636   .3644382    -0.07   0.947    -.7385494    .6900222
-------------+----------------------------------------------------------------
       _cut1 |  -2.065609   .5583871          (Ancillary parameters)
       _cut2 |  -1.539941   .5510676
       _cut3 |  -1.309679   .5479021
       _cut4 |   -.665663    .543653
       _cut5 |  -.5036779   .5442082
------------------------------------------------------------------------------
```

Marginal effects:

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

Marginal effects after oprobit
      y  = Pr(obribe1==1) (predict, outcome(1))
         =  .26194813
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      lk |   .0263548      .01007    2.62   0.009   .006619   .04609    15.675
   profk |   .0185525      .01232    1.51   0.132  -.005599  .042704  -.364765
    sec2*|   .1920139      .10092    1.90   0.057  -.005785  .389813        .2
    sec3*|  -.0342657      .07595   -0.45   0.652  -.183127  .114596   .290323
    sec4*|   .1359843      .09025    1.51   0.132  -.040909  .312877   .251613
 nairobi*|  -.0241228      .09275   -0.26   0.795  -.205901  .157656   .509677
 mombasa*|   .0479847      .10129    0.47   0.636  -.150547  .246517   .264516
  nakuru*|   .0079487      .12011    0.07   0.947   -.22747  .243368   .103226
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

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

Marginal effects after oprobit
      y  = Pr(obribe1==3) (predict, outcome(3))
         =  .09165806
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      lk |  -.0000255      .00073   -0.03   0.972  -.001459  .001408    15.675
   profk |  -.0000179      .00051   -0.03   0.972  -.001027  .000991  -.364765
    sec2*|  -.0078598      .00883   -0.89   0.374  -.025173  .009453        .2
    sec3*|  -.0001844      .00132   -0.14   0.889  -.002777  .002408   .290323
    sec4*|  -.0035893      .00571   -0.63   0.529  -.014771  .007593   .251613
 nairobi*|   .0000281      .00068    0.04   0.967  -.001302  .001358   .509677
 mombasa*|  -.0004917      .00236   -0.21   0.835  -.005126  .004143   .264516
  nakuru*|  -.0000289      .00079   -0.04   0.971   -.00158  .001522   .103226
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

. mfx compute, predict(outcome(6));

Marginal effects after oprobit
      y  = Pr(obribe1==6) (predict, outcome(6))
         =  .17759222
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      lk |  -.0210592      .00817   -2.58   0.010   -.03708 -.005038    15.675
   profk |  -.0148246       .0099   -1.50   0.134  -.034227  .004578  -.364765
    sec2*|  -.1203152        .051   -2.36   0.018  -.220279 -.020351        .2
    sec3*|   .0283602      .06519    0.44   0.664  -.099409  .156129   .290323
    sec4*|  -.0936442      .05426   -1.73   0.084  -.199983  .012695   .251613
 nairobi*|   .0192561      .07374    0.26   0.794   -.12528  .163792   .509677
 mombasa*|  -.0363774      .07328   -0.50   0.620  -.179994  .107239   .264516
  nakuru*|  -.0062568      .09313   -0.07   0.946  -.188796  .176283   .103226
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Note: The sign of the marginal effects referring to the highest outcome are the same as the sign of the estimated parameter beta(j), and the sign of the marginal effects referring to the lowest outcome are the opposite to the sign of the estimated parameter beta(j). For intermediate outcome categories, the signs of the marginal effects are ambiguous and often close to zero (e.g. outcome 3 above). Why is this?

2. Multinomial Logit

In the following example we consider a model of occupational choice within the Kenyan manufacturing sector (see footnote 1 for a reference for the data). We have data on 950 individuals and we want to investigate if education, gender and parental background determine occupation.

We distinguish between four classes of jobs:

- management
- administration and supervision
- sales and support staff
- production workers

Sample proportions for these four categories are as follows:

```
. tabulate job

        job |      Freq.     Percent        Cum.
------------+-----------------------------------
       Prod |        545       57.37       57.37
      Manag |         91        9.58       66.95
      Admin |        270       28.42       95.37
    Support |         44        4.63      100.00
------------+-----------------------------------
      Total |        950      100.00
```

The explanatory variables are

years of education: educ
gender: male
parental background: f_prof, m_prof (father/mother professional), f_se, m_se (father/mother self-employed or trader)

Summary statistics for these variables are as follows:

```
. sum educ male f_prof f_se m_prof m_se;

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        educ |       950    9.933684     2.86228          0         17
        male |       950    .8136842    .3895664          0          1
      f_prof |       950    .1347368    .3416221          0          1
        f_se |       950    .1231579    .3287915          0          1
      m_prof |       950    .0578947    .2336673          0          1
-------------+--------------------------------------------------------
        m_se |       950    .1315789    .3382105          0          1
```

A breakdown by occupation is a useful first step to see if there are any broad patterns in the data:

```
. tabstat educ male f_prof f_se m_prof m_se, by(job);

Summary statistics: mean
  by categories of: job

     job |      educ      male    f_prof      f_se    m_prof      m_se
---------+------------------------------------------------------------
    Prod |  8.946789  .8825688  .0715596  .1229358  .0348624  .1559633
   Manag |  12.82418  .8021978  .3406593  .1208791  .1318681  .0879121
   Admin |  10.75926  .7037037  .1814815  .1074074  .0666667        .1
 Support |  11.11364  .6590909  .2045455  .2272727  .1363636  .1136364
---------+------------------------------------------------------------
   Total |  9.933684  .8136842  .1347368  .1231579  .0578947  .1315789
----------------------------------------------------------------------
```

The multinomial logit seems a suitable model for modelling occupational choice with these data (notice in particular that there is no natural ordering of the dependent variable).

I begin by coding the job variable from 0 to 3:

job: 0 = prod; 1 = manag; 2 = admin; 3 = support

So I will obtain three vectors of parameter estimates. Because I have set job = 0 for production workers, this will be the **base category** (I can alter this by using the *basecategory( )* option).

Results:

```
. mlogit job educ male f_prof f_se m_prof m_se;

Multinomial logistic regression                   Number of obs   =        950
                                                  LR chi2(18)     =     289.97
                                                  Prob > chi2     =     0.0000
Log likelihood = -846.16161                       Pseudo R2       =     0.1463

------------------------------------------------------------------------------
         job |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Manag        |
        educ |    .738846   .0755869     9.77   0.000     .5906984    .8869935
        male |   .0277387   .3383262     0.08   0.935    -.6353685     .690846
      f_prof |   1.135737   .3373116     3.37   0.001     .4746187    1.796856
        f_se |   .1189543   .4074929     0.29   0.770     -.679717    .9176256
      m_prof |   .3806786   .4661837     0.82   0.414    -.5330247    1.294382
        m_se |  -.6073577   .4413568    -1.38   0.169    -1.472401    .2576856
       _cons |  -10.25324   .9913425   -10.34   0.000    -12.19623   -8.310244
-------------+----------------------------------------------------------------
Admin        |
        educ |   .2421636   .0333887     7.25   0.000     .1767229    .3076042
        male |  -.9075081   .2018354    -4.50   0.000    -1.303098    -.511918
      f_prof |   .5696015   .2570499     2.22   0.027      .065793     1.07341
        f_se |  -.0884656   .2616688    -0.34   0.735     -.601327    .4243958
      m_prof |  -.0135092   .3751632    -0.04   0.971    -.7488156    .7217972
        m_se |  -.5700617    .256966    -2.22   0.027    -1.073706   -.0664175
       _cons |  -2.350944   .3941898    -5.96   0.000    -3.123542   -1.578346
-------------+----------------------------------------------------------------
Support      |
        educ |   .2805316   .0723475     3.88   0.000     .1387331    .4223302
        male |  -.9905816   .3642871    -2.72   0.007    -1.704571    -.276592
      f_prof |   .6547286   .4707312     1.39   0.164    -.2678877    1.577345
        f_se |   .8717071   .4237441     2.06   0.040     .0411839     1.70223
      m_prof |   .7996763   .5500412     1.45   0.146    -.2783846    1.877737
        m_se |  -.5924061   .5213599    -1.14   0.256    -1.614253    .4294405
       _cons |  -4.777905   .8675103    -5.51   0.000    -6.478193   -3.077616
------------------------------------------------------------------------------
(Outcome job==Prod is the comparison group)
```

Marginal effects

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

Marginal effects after mlogit
      y  = Pr(job==1) (predict, outcome(1))
         =  .03792548
-------------------------------------------------------------------------------
                       variable |          dy/dx                 X
---------------------------------+---------------------------------------------
                           educ |       .0236809           9.93368
                          male*|       .0136163           .813684
                        f_prof*|       .0448291           .134737
                          f_se*|       .0033605           .123158
                        m_prof*|       .0139252           .057895
                          m_se*|      -.0134499           .131579
-------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

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

Marginal effects after mlogit
      y  = Pr(job==2) (predict, outcome(2))
         =  .30591218
-------------------------------------------------------------------------------
                        variable |          dy/dx                 X
---------------------------------+---------------------------------------------
                            educ |        .0390723             9.93368
                            male*|       -.1879454             .813684
                          f_prof*|        .0950757             .134737
                            f_se*|       -.0354302             .123158
                          m_prof*|       -.0225145             .057895
                            m_se*|       -.0999013             .131579
-------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

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

Marginal effects after mlogit
      y  = Pr(job==3) (predict, outcome(3))
         =  .04398022
-------------------------------------------------------------------------------
                        variable |          dy/dx                 X
---------------------------------+---------------------------------------------
                            educ |        .0073048             9.93368
                            male*|       -.0322613             .813684
                          f_prof*|        .0184086             .134737
                            f_se*|        .0519705             .123158
                          m_prof*|        .0458739             .057895
                            m_se*|        -.015106             .131579
-------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

. mfx compute, predict(outcome(0)) nose;

Marginal effects after mlogit
      y  = Pr(job==0) (predict, outcome(0))
         =  .61218213
-------------------------------------------------------------------------------
                        variable |          dy/dx                 X
---------------------------------+---------------------------------------------
                            educ |       -.0700579             9.93368
                            male*|        .2065904             .813684
                          f_prof*|       -.1583134             .134737
                            f_se*|       -.0199008             .123158
                          m_prof*|       -.0372846             .057895
                            m_se*|        .1284572             .131579
-------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Predicted job probabilities

Education = PRIMARY, SECONDARY, and UNIVERSITY

```
1. PRIMARY
. list pp1 pp2 pp3 pp0;

     +-------------------------------------------+
     |      pp1        pp2        pp3        pp0 |
     |-------------------------------------------|
  1. | .0108399   .2284537   .0304956   .7302108 |
     +-------------------------------------------+

2. SECONDARY
. list sp1 sp2 sp3 sp0;

     +-------------------------------------------+
     |      sp1        sp2        sp3        sp0 |
     |-------------------------------------------|
  1. | .1274372   .3683361   .0573239   .4469028 |
     +-------------------------------------------+

3. UNIVERSITY

. list up1 up2 up3 up0;

     +-----------------------------------------+
     |      up1       up2        up3       up0 |
     |-----------------------------------------|
  1. | .6057396   .240109   .0435665   .110585 |
     +-----------------------------------------+

Note: 1 = manag; 2 = admin; 3 = support; 0 = prod
```

<u>How these probabilities were calculated:</u>

```
/* first collapse the data: this gives a new data set consisting of one
observations and the sample means of the variables */

. collapse educ male f_prof f_se m_prof m_se;

/* now vary education: since 1985 the Kenyan education system has involved 8 years
for primary education, 4 years for secondary, and 4 years for university */

/* first do primary */
. replace educ = 8;
(1 real change made)

/* get the predicted probability that the ‘individual’ is a manager */
> predict pp1, outcome(1);
(option p assumed; predicted probability)

/* get the predicted probability that the ‘individual’ is admin */
. predict pp2, outcome(2);
(option p assumed; predicted probability)

. predict pp3, outcome(3);
(option p assumed; predicted probability)

. predict pp0, outcome(0);
(option p assumed; predicted probability)

/* now do secondary */
. replace educ=12;
(1 real change made)
```

```
>
> predict sp1, outcome(1);
(option p assumed; predicted probability)

. predict sp2, outcome(2);
(option p assumed; predicted probability)

. predict sp3, outcome(3);
(option p assumed; predicted probability)

. predict sp0, outcome(0);
(option p assumed; predicted probability)

/* finally do university */
. replace educ=16;
(1 real change made)


> predict up1, outcome(1);
(option p assumed; predicted probability)

. predict up2, outcome(2);
(option p assumed; predicted probability)

. predict up3, outcome(3);
(option p assumed; predicted probability)

. predict up0, outcome(0);
(option p assumed; predicted probability)
```

3. Illustration: The Hausman test for IIA in multinomial logit is totally useless

```
. use http://www.stata-press.com/data/r9/sysdsn3, clear

(Health insurance data)

.
. /* The results shown in the manual */
. mlogit insure male age

Iteration 0:   log likelihood = -555.85446
Iteration 1:   log likelihood = -551.32973
Iteration 2:   log likelihood = -551.32802

Multinomial logistic regression                   Number of obs   =        615
                                                  LR chi2(4)      =       9.05
                                                  Prob > chi2     =     0.0598
Log likelihood = -551.32802                       Pseudo R2       =     0.0081

------------------------------------------------------------------------------
      insure |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Prepaid      |
        male |   .5095747   .1977893     2.58   0.010     .1219148    .8972345
         age |  -.0100251   .0060181    -1.67   0.096    -.0218204    .0017702
       _cons |   .2633838   .2787574     0.94   0.345    -.2829708    .8097383
-------------+----------------------------------------------------------------
Uninsure     |
        male |   .4748547   .3618446     1.31   0.189    -.2343477    1.184057
         age |  -.0051925   .0113821    -0.46   0.648     -.027501     .017116
       _cons |  -1.756843   .5309591    -3.31   0.001    -2.797504   -.7161824
------------------------------------------------------------------------------
(insure==Indemnity is the base outcome)

. estimates store allcats

.
. mlogit insure male age if insure !="Uninsure":insure

Iteration 0:   log likelihood =  -394.8693
Iteration 1:   log likelihood =  -390.4871
Iteration 2:   log likelihood = -390.48643

Multinomial logistic regression                   Number of obs   =        570
                                                  LR chi2(2)      =       8.77
                                                  Prob > chi2     =     0.0125
Log likelihood = -390.48643                       Pseudo R2       =     0.0111

------------------------------------------------------------------------------
      insure |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Prepaid      |
        male |   .5144003   .1981735     2.60   0.009     .1259875    .9028132
         age |  -.0101521   .0060049    -1.69   0.091    -.0219214    .0016173
       _cons |   .2678043   .2775562     0.96   0.335    -.2761959    .8118046
------------------------------------------------------------------------------
(insure==Indemnity is the base outcome)

.
```

```
. hausman . allcats, alleqs constant

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       .         allcats       Difference          S.E.
-------------+----------------------------------------------------------------
        male |    .5144003     .5095747        .0048256         .012334
         age |   -.0101521    -.0100251       -.0001269               .
       _cons |    .2678043     .2633838        .0044205               .
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from mlogit
           B = inconsistent under Ha, efficient under Ho; obtained from mlogit

    Test:  Ho:  difference in coefficients not systematic

                  chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        0.08
                Prob>chi2 =      0.9944
                (V_b-V_B is not positive definite)

.
. /* confirm that IIA test is nonsense in model with male dummy only */
. mlogit insure male

Iteration 0:   log likelihood = -556.59502
Iteration 1:   log likelihood = -553.40794
Iteration 2:   log likelihood = -553.40712

Multinomial logistic regression                   Number of obs   =        616
                                                  LR chi2(2)      =       6.38
                                                  Prob > chi2     =     0.0413
Log likelihood = -553.40712                       Pseudo R2       =     0.0057

------------------------------------------------------------------------------
      insure |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Prepaid      |
        male |    .477311   .1959282     2.44   0.015     .0932987    .8613234
       _cons |  -.1772065   .0968274    -1.83   0.067    -.3669847    .0125718
-------------+----------------------------------------------------------------
Uninsure     |
        male |     .46019   .3593228     1.28   0.200    -.2440698     1.16445
       _cons |  -1.989585   .1884768   -10.56   0.000    -2.358993   -1.620177
------------------------------------------------------------------------------
(insure==Indemnity is the base outcome)

. estimates store allcats

.
. mlogit insure male if insure !="Uninsure":insure

Iteration 0:   log likelihood = -395.53394
Iteration 1:   log likelihood = -392.53619
Iteration 2:   log likelihood = -392.53611

Multinomial logistic regression                   Number of obs   =        571
                                                  LR chi2(1)      =       6.00
                                                  Prob > chi2     =     0.0143
Log likelihood = -392.53611                       Pseudo R2       =     0.0076

------------------------------------------------------------------------------
      insure |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Prepaid      |
        male |    .477311   .1959283     2.44   0.015     .0932987    .8613234
       _cons |  -.1772065   .0968274    -1.83   0.067    -.3669847    .0125718
------------------------------------------------------------------------------
(insure==Indemnity is the base outcome)
```

```
.
. hausman . allcats, alleqs constant

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       .         allcats       Difference          S.E.
-------------+----------------------------------------------------------------
        male |     .477311      .477311        2.63e-13         .000109
       _cons |   -.1772065    -.1772065       -3.66e-15               .
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from mlogit
           B = inconsistent under Ha, efficient under Ho; obtained from mlogit

    Test:  Ho:  difference in coefficients not systematic

                  chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        0.00
                Prob>chi2 =      1.0000
                (V_b-V_B is not positive definite)

.
. /* confirm that IIA test is nonsense in model with constant only */
. mlogit insure

Iteration 0:   log likelihood = -556.59502

Multinomial logistic regression                   Number of obs   =        616
                                                  LR chi2(0)      =       0.00
                                                  Prob > chi2     =          .
Log likelihood = -556.59502                       Pseudo R2       =     0.0000

------------------------------------------------------------------------------
      insure |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Prepaid      |
       _cons |  -.0595623   .0837345    -0.71   0.477    -.2236789    .1045544
-------------+----------------------------------------------------------------
Uninsure     |
       _cons |  -1.876917   .1600737   -11.73   0.000    -2.190656   -1.563179
------------------------------------------------------------------------------
(insure==Indemnity is the base outcome)

. estimates store allcats

.
. mlogit insure if insure !="Uninsure":insure

Iteration 0:   log likelihood = -395.53394

Multinomial logistic regression                   Number of obs   =        571
                                                  LR chi2(0)      =       0.00
                                                  Prob > chi2     =          .
Log likelihood = -395.53394                       Pseudo R2       =     0.0000

------------------------------------------------------------------------------
      insure |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Prepaid      |
       _cons |  -.0595623   .0837345    -0.71   0.477    -.2236789    .1045544
------------------------------------------------------------------------------
(insure==Indemnity is the base outcome)

.
```

```
. hausman . allcats, alleqs constant

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       .         allcats       Difference          S.E.
-------------+----------------------------------------------------------------
       _cons |   -.0595623    -.0595623        7.69e-15               .
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from mlogit
           B = inconsistent under Ha, efficient under Ho; obtained from mlogit

    Test:  Ho:  difference in coefficients not systematic

                  chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =    -0.00    chi2<0 ==> model fitted on these
                                        data fails to meet the asymptotic
                                        assumptions of the Hausman test;
                                        see suest for a generalized test
```

4. Tobit

Figure 1. The inverse Mills ratio function

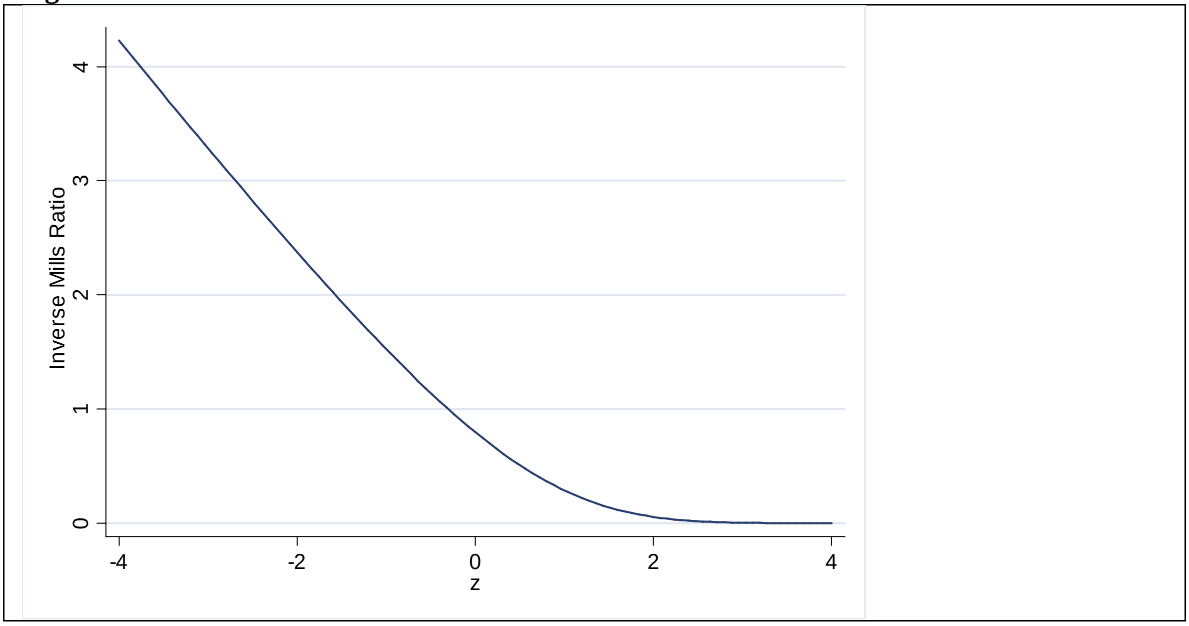


Illustration: Modelling investment among Ghanaian manufacturing firms

In the following example we consider a model of company investment within the Ghanaian manufacturing sector.[2] Our dataset consists of 1,202 observations on firms over the 1991-99 period (in fact, there is a panel dimension in the data, but we will ignore this for now).

Our simple model of investment is

$$\left(\frac{I}{K}\right)_{it} = \max\{0, \alpha_0 + \alpha_1 \ln TFP_{it} + \alpha_2 \ln K_{i,t-1} + u_{it}\},$$

where

I = Gross investment in fixed capital (plant & machinery)
K = Value of the capital stock
TFP = Total factor productivity, defined as ln(output) – 0.3ln(K) – 0.7ln(L), where L is employment
u = a residual, assumed homoskedastic and normally distributed.

There is evidence physical capital is 'irreversible' in African manufacturing, i.e. selling off fixed capital is difficult due to the lack of a market for second hand capital goods (Bigsten et al., 2005). We can thus view investment as a corner response variable: investment is bounded below at zero.

Summary statistics for these variables are as follows:

[2] This is an extension of the dataset used by Söderbom and Teal (2004).

Table 1. Summary statistics

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     invrate |      1202    .0629597    .1477861          0          1
      invdum |      1202    .4550749    .4981849          0          1
         tfp |      1202    10.20903    1.108122   5.049412    14.7326
        lk_1 |      1202    16.06473    3.104121   9.555573   23.51505
```

Note: invrate = (I/K); invdum = 1 if invrate>0, = 0 if invrate=0; tfp = ln(TFP);
lk_1 = ln[K(t-1)]

Table 2. OLS results

```
. reg invrate tfp lk_1;

      Source |       SS       df       MS              Number of obs =    1202
-------------+------------------------------           F(  2,  1199) =    4.54
       Model |  .197262981     2  .098631491           Prob > F      =  0.0108
    Residual |  26.0334412  1199  .021712628           R-squared     =  0.0075
-------------+------------------------------           Adj R-squared =  0.0059
       Total |  26.2307042  1201   .02184072           Root MSE      =  .14735

------------------------------------------------------------------------------
     invrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         tfp |   .0114908   .0038443     2.99   0.003     .0039484    .0190331
        lk_1 |   .0002798   .0013724     0.20   0.838    -.0024127    .0029723
       _cons |   -.058845   .0440225    -1.34   0.182    -.1452148    .0275248
------------------------------------------------------------------------------
```

Table 3. Tobit results

```
. tobit invrate tfp lk_1, ll(0);

Tobit estimates                                   Number of obs   =       1202
                                                  LR chi2(2)      =      44.34
                                                  Prob > chi2     =     0.0000
Log likelihood =  -398.5866                       Pseudo R2       =     0.0527

------------------------------------------------------------------------------
     invrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         tfp |   .0344135   .0077922     4.42   0.000     .0191257    .0497012
        lk_1 |   .0123672   .0027384     4.52   0.000     .0069947    .0177397
       _cons |  -.6158372   .0913444    -6.74   0.000    -.7950496   -.4366247
-------------+----------------------------------------------------------------
         _se |   .2540915   .0083427           (Ancillary parameter)
------------------------------------------------------------------------------

  Obs. summary:        655  left-censored observations at invrate<=0
                       547     uncensored observations
```

Marginal effects based on tobit

```
. mfx compute, predict(e(0,.));

Marginal effects after tobit
      y  = E(invrate|invrate>0) (predict, e(0,.))
         =  .18058807
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
     tfp |   .0106934      .00241    4.44   0.000   .005969  .015418    10.209
    lk_1 |   .0038429      .00084    4.56   0.000    .00219  .005496   16.0647
------------------------------------------------------------------------------
```

Table 4. Probit modelling whether or not firm invests at all

```
Probit estimates                                  Number of obs   =       1202
                                                  LR chi2(2)      =     116.98
                                                  Prob > chi2     =     0.0000
Log likelihood =  -769.8164                       Pseudo R2       =     0.0706

------------------------------------------------------------------------------
      invdum |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         tfp |   .1777798   .0345322     5.15   0.000     .1100979    .2454616
        lk_1 |    .112731   .0123721     9.11   0.000     .0884821      .13698
       _cons |  -3.744003   .4070374    -9.20   0.000    -4.541782   -2.946225
------------------------------------------------------------------------------
```

Marginal effects after probit

```
      y  = Pr(invdum) (predict)
         =  .45301349
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
     tfp |   .0704314      .01368    5.15   0.000   .043617  .097245   10.209
    lk_1 |   .0446609       .0049    9.11   0.000    .03505  .054272  16.0647
------------------------------------------------------------------------------
```

Testing the 'single mechanism' assumption

In both models, the estimated coefficients on tfp and lk_1 are positive and significant, suggesting the 'single mechanism' assumption is OK. Now look at the point estimates:

Check if

```
i)          _b[tfp](tobit)/sdev(tobit) ~= _b[tfp](probit)
. disp .0344135/.2540915
.13543743
```

which is reasonably close to the probit estimate of 0.178 (at least it doesn't look significantly different)

```
ii)         _b[lk_1](tobit)/sdev(tobit) ~= _b[lk_1](probit)

. disp  .0123672 /.2540915
.04867223
```

which is rather much lower than the probit estimate of 0.113, suggesting the tobit specification may be problematic. This deserves further investigation.

Functional form test for tobit

```
. ge xb2=xb^2;

. ge xb3=xb^3;
```

Table 5: Alternative tobit model, containing a nonlinear function of xb

```
. tobit invrate xb2 xb3, offset(xb) ll(0) nocons;

Tobit estimates                                   Number of obs   =       1202
                                                  LR chi2(1)      =          .
Log likelihood = -398.40664                       Prob > chi2     =          .

------------------------------------------------------------------------------
     invrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         xb2 |   .0897647   2.674486     0.03   0.973     -5.15742    5.336949
         xb3 |  -2.696379   16.09553    -0.17   0.867    -34.27487    28.88211
          xb |   (offset)
-------------+----------------------------------------------------------------
         _se |   .2524378    .008011          (Ancillary parameter)
------------------------------------------------------------------------------

  Obs. summary:        655  left-censored observations at invrate<=0
                       547     uncensored observations
```

Notice the two options used here (the instructions after the , in tobit): offset(xb) is equivalent to imposing a coefficient equal to one on xb, and nocons means the model is estimated without a constant.

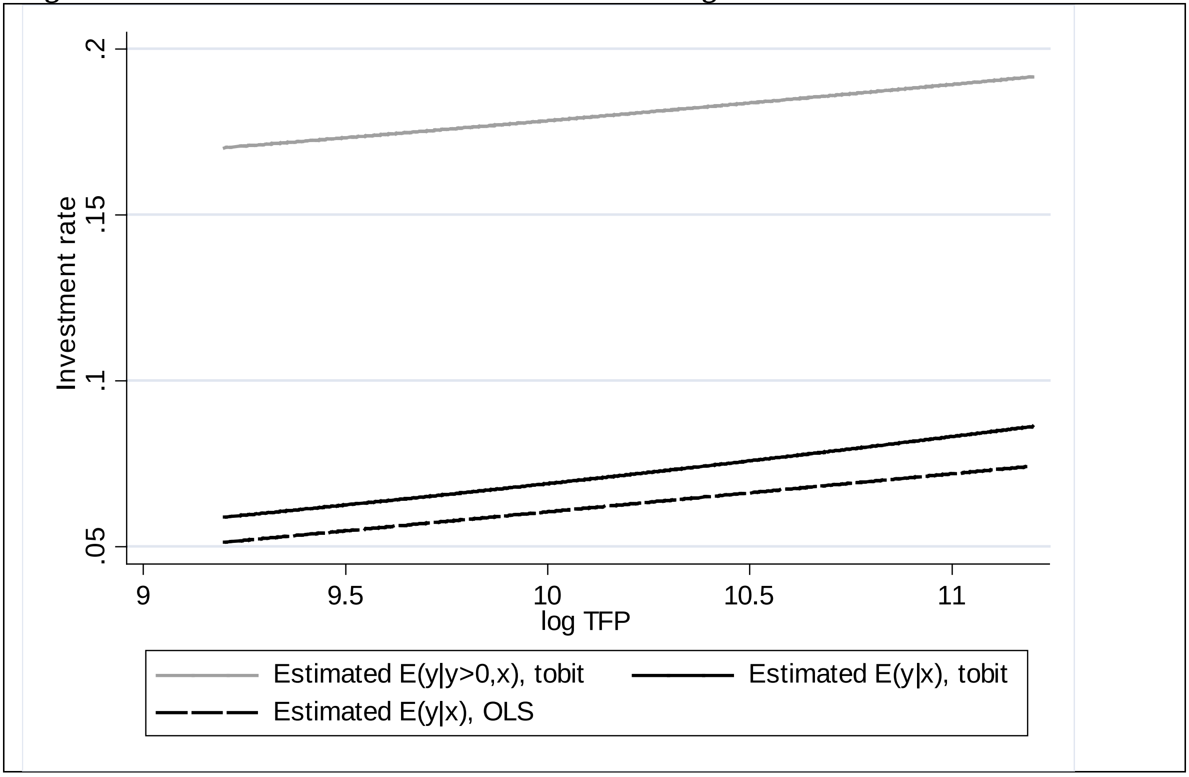
Wald test:

```
. test xb2 xb3

 ( 1)  xb2 = 0
 ( 2)  xb3 = 0

       F(  2,  1201) =    0.18
            Prob > F =    0.8317
```

Figure 2. Predicted investment rates as a function of log TFP



Note: Evaluated at the sample mean of lk_1.

5. Illustration: The truncated regression model

Consider a simple simulation, obtained by the following Stata code:

```
clear
set seed 2355
set obs 500

ge u=invnorm(uniform())

ge x=2*uniform()

/* true population model: y = -1 + 1*x + u /

ge y=-1+x+u

/* no truncation */
reg y x
predict yh_ols_nt

/* truncation of y at 0.8*/
reg y x if y<.8
predict yh_ols_t

/* truncated regression corrects for the trunctation. ul(.) indicates the upper limit */

truncreg y x, ul(0.8)
```

Consider three different regressions based on these artificial data:

i) OLS using the full sample of 500 observations (i.e. no truncation)

```
. reg y x

      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  1,   498) =  156.47
       Model |  139.883218     1  139.883218           Prob > F      =  0.0000
    Residual |  445.219899   498  .894015862           R-squared     =  0.2391
-------------+------------------------------           Adj R-squared =  0.2375
       Total |  585.103118   499  1.17255134           Root MSE      =  .94552

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .8940591   .0714753    12.51   0.000     .7536288    1.034489
       _cons |  -.9019037   .0834538   -10.81   0.000    -1.065869   -.7379389
------------------------------------------------------------------------------
```

ii) OLS using the truncated sample of 380 observations

```
. reg y x if y<.8

      Source |       SS       df       MS              Number of obs =     380
-------------+------------------------------           F(  1,   378) =   47.00
       Model |   28.616886     1   28.616886           Prob > F      =  0.0000
    Residual |  230.164146   378  .608899857           R-squared     =  0.1106
-------------+------------------------------           Adj R-squared =  0.1082
       Total |  258.781032   379  .682799556           Root MSE      =  .78032

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .4811388    .070183     6.86   0.000     .3431407    .6191369
       _cons |  -.8577185   .0732374   -11.71   0.000    -1.001722   -.7137147
------------------------------------------------------------------------------
```

Notice coefficient on x is much lower than the true value of one. It is clearly significantly different from one, indicating significant bias.

Figure 3 illustrates the problem of truncation.

iii) Truncated regression which corrects for the truncation

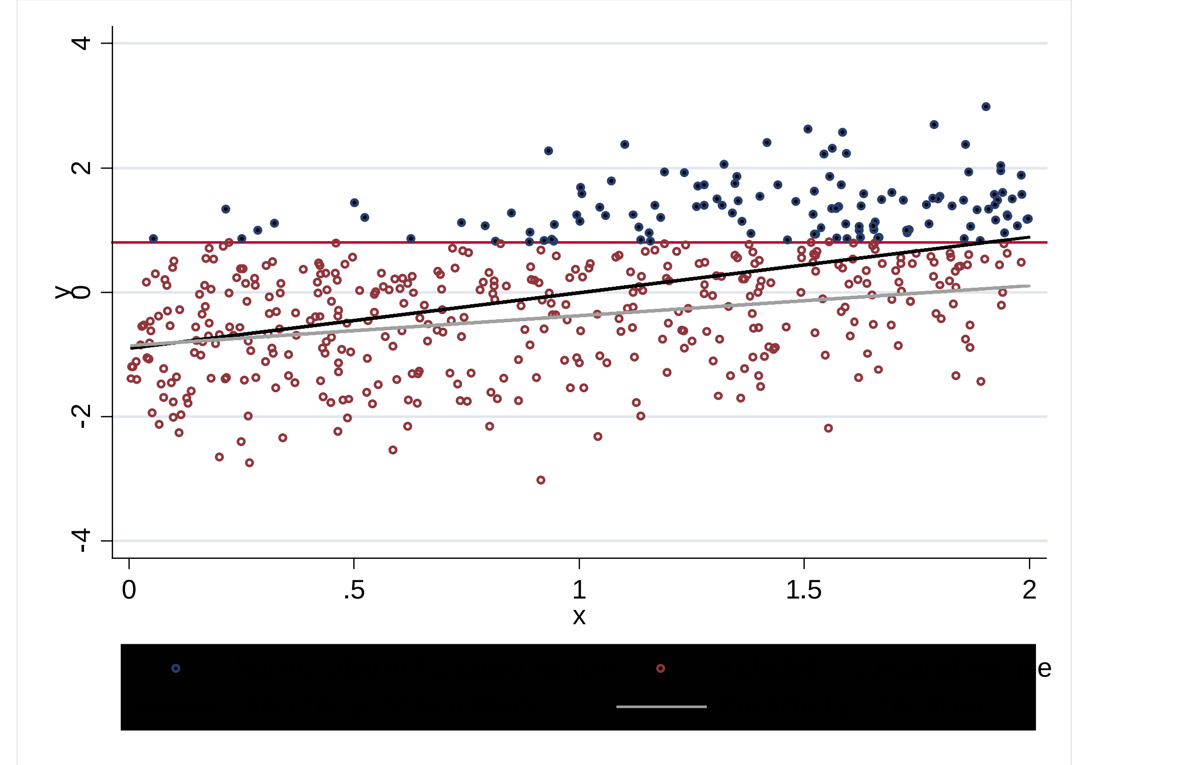
```
. truncreg y x, ul(0.8)
(note: 120 obs. truncated)

Truncated regression
Limit:   lower =       -inf                             Number of obs =    380
         upper =         .8                             Wald chi2(1)  =  37.41
Log likelihood = -398.51329                             Prob > chi2   = 0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
eq1          |
           x |   .8506762   .1390748     6.12   0.000     .5780947    1.123258
       _cons |  -.7836381   .1214471    -6.45   0.000     -1.02167   -.5456061
-------------+----------------------------------------------------------------
sigma        |
       _cons |   1.019341    .067624    15.07   0.000     .8868003    1.151882
------------------------------------------------------------------------------
```

Coefficient increases as a result and is similar to the OLS estimate in (i) and not significantly different from the true value of 1.

Figure 3. The effect of truncation on the OLS estimator



Note: The predications have been generated from the OLS estimates shown in (i) and (ii) above.

References

Bigsten, Arne, Paul Collier, Stefan Dercon, Marcel Fafchamps, Bernard Gauthier, Jan Willem Gunning, Remco Oostendorp, Catherine Pattillo, Måns Söderbom, and Francis Teal (2005). “Adjustment Costs, Irreversibility and Investment Patterns in African Manufacturing,” *The B.E. Journals in Economic Analysis & Policy: Contributions to Economic Analysis & Policy* 4:1, Article 12, pp. 1-27.

Söderbom, Måns, and Francis Teal (2004). “Size and Efficiency in African Manufacturing Firms: Evidence from Firm-Level Panel Data,” *Journal of Development Economics* 73, pp. 369-394.