Barcelona GSE, Universitat Pompeu Fabra
Empirical Methods for Applied Analysis

Kurt Schmidheiny
November 2008

Limited Dependent Variable Models

Contents

1 Truncation

Truncation occurs when the observed data in the sample are drawn from a subset of the whole population. The subset is defined based on the value of the *dependent* variable.

An example: A study of the determinants of incomes of the poor. Only households with income below a certain poverty line are part of the sample.

1.1 The Model (Truncated Regression)

Consider a latent random variable y_i that linearly depends on x_i, i.e.

$$y_i^* = x_i'\beta + \varepsilon_i \quad \text{with} \quad \varepsilon_i|x_i \sim N(0,\sigma^2).$$

The error term ε_i is independently and normally distributed with mean 0 and variance σ^2. The distribution of y_i^* given x_i is therefore also normal: $y_i^*|x_i \sim N(x_i'\beta,\sigma^2)$. The expected value of the latent variable is $E(y_i^*|x_i) = x_i'\beta$.

Observation i is only observed if y_i^* is above a certain known threshold a, i.e.

$$y_i = \begin{cases} y_i^* & \text{if} \quad y_i^* > a \\ \text{n.a.} & \text{if} \quad y_i^* \leq a \end{cases}$$

The density function of the observed truncated variable y_i is therefore the pdf of the latent variable conditional on it being observed, i.e.[1]

$$f(y_i|x_i) = f\left(y_i^*\,|y_i^* > a, x_i\right) = \frac{f(y_i^*|x_i)}{P(y_i^* > a|x_i)} = \frac{\sigma^{-1}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)}{1-\Phi\left(\frac{a-x_i'\beta}{\sigma}\right)} = \frac{1}{\sigma}\frac{\phi\left(\frac{x_i'\beta - y_i}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta - a}{\sigma}\right)}$$

[1]Note how the pdf of a normally distributed variable ε with mean μ variance σ^2 can be written using the pdf $\phi(.)$ of the standard normal $N(0,1)$

$$f(\varepsilon) = \frac{1}{\sigma\sqrt{2\pi}}\, exp\left[-\frac{(\varepsilon-\mu)^2}{2\sigma^2}\right] = \frac{1}{\sigma}\left\{\frac{1}{\sqrt{2\pi}}\, exp\left[-\frac{1}{2}\left(\frac{\varepsilon-\mu}{\sigma}\right)^2\right]\right\} = \sigma^{-1}\phi\left(\frac{\varepsilon-\mu}{\sigma}\right)$$

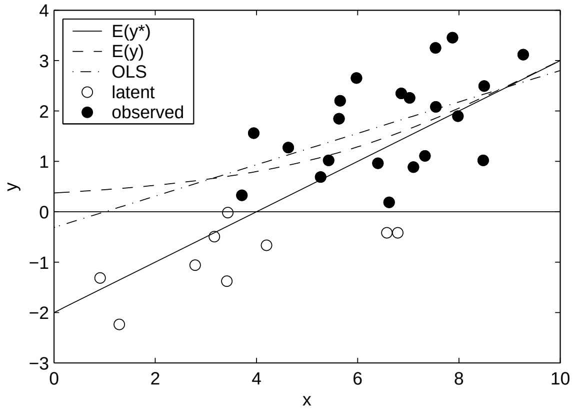

Figure 1: The truncated regression model. Lower truncation at $a = 0$, $N = 30$, $K = 2$, $\beta = (-2, 0.5)'$ and $\sigma = 1$.

where $\phi(.)$ is the pdf and $\Phi(.)$ the cumulative normal distribution.

The expected value of the observed variable is not linear in x_i (try to derive the equation below)

$$E(y_i|x_i) = E\left(y_i^* \, | y_i^* > a, x_i\right) = x_i'\beta + \sigma \frac{\phi\left[(x_i'\beta - a)/\sigma\right]}{\Phi\left[(x_i'\beta - a)/\sigma\right]} = x_i'\beta + \sigma\lambda_i$$

where $\lambda_i \equiv \phi(\alpha_i)/\Phi(\alpha_i)$ and $\alpha_i = (x_i'\beta - a)/\sigma$. Figure 1 visualizes the truncated regression model in an example.

1.2 Interpretation of Parameters

The interpretation of the parameters depends very much on the research question. If the researcher is interested in the underlying linear relationship in the whole population, the slope coefficients β can simply be interpreted as marginal effects. However, if the researcher is only interested in the effect on the observed subpopulation, the marginal effect is

more complicated:

$$\begin{aligned} \frac{\partial E(y_i|x_i)}{\partial x_{ik}} &= \frac{\partial E\left(y_i^* \, | y_i^* > a, x_i\right)}{\partial x_{ik}} \\ &= \beta_k + \sigma \frac{\partial \lambda_i}{\partial x_{ik}} = \beta_k \left[1 - \lambda_i^2 - \alpha_i \lambda_i\right]. \end{aligned}$$

1.3 Estimation

OLS regression of the observed variable y_i on x_i

$$y_i = x_i'\beta + u_i$$

will yield biased estimates of beta, as the error term $u_i = \varepsilon_i | y_i^* > a$ is correlated with x_i and $E(u_i|x_i) = E(\varepsilon_i|y_i^* > a, x_i) = \sigma\lambda_i > 0$.

The truncated regression is therefore usually estimated by maximum likelihood (ML). The log likelihood function is

$$\ln \mathcal{L} = \sum_{i=1}^{N} \ln \left[\sigma^{-1} \phi \left(\frac{y_i - x_i'\beta}{\sigma}\right)\right] - \sum_{i=1}^{N} \ln \left[1 - \Phi\left(\frac{a - x_i'\beta}{\sigma}\right)\right]$$

and allows to estimate both β and σ by an iterative numerical procedure. Usual ML properties (consistency, asymptotic efficiency and normality, etc) apply and asymptotic hypothesis tests can be performed as Wald, likelihood ratio or lagrange multiplier tests.

1.4 Implementation in STATA 10.0

Stata estimates the truncated regression model by the command

```
truncreg depvar [indepvars], ll(#)
```

where `ll(#)` defines the lower truncation point a. We can also estimate a more general model with a lower and an upper truncation point

```
truncreg depvar [indepvars], ll(varname) lu(varname)
```

where the upper `ll` and lower `lu` thresholds can be observation specific and their values are defined by *varname*.

The post-estimation commands `predict` calculates by default the linear index function $x_i'\beta$. The option `predict, e(a,b)` calculates $E(y_i|x_i) = E(y_i^*|a < y_i^* < b, x_i)$. The post-estimation command `mfx` reports marginal effects. By default $\partial E(y_i^*|x_i)/\partial x_{ik} = \beta_k$ is reported. The marginal effect on the observed truncated variable, $\partial E(y_i^*|a < y_i^* < b, x_i)/\partial x_{ik}$, is calculated by `mfx, predict(e(a,b))`.

For example,

```
webuse laborsub
truncreg whrs kl6 k618 wa we, ll(0)
predict whrs_hat, e(0,.)
mfx, predict(e(0,.)) at(kl6=1, k618=0, wa=35, we=15)
```

regresses the hours worked (*whrs*) on the number of children below age 6 (*kl6*), children between age 6 and 18 (*k618*), age (*wa*) and education in years (*we*) for a sample women who work (*whrs*> 0). It then predicts the hours worked in the subpopulation of working women, i.e. $E(whrs_i|x_i) = E(whrs_i^*|whrs_i^* > 0, x_i)$. Marginal effects on the hours worked in the subpopulation, i.e. $\partial E(whrs_i|x_i)/\partial x_{ik}$, are then reported for a 35 year old woman with 15 years of education and one child below age 6.

2 Censoring

Censoring occurs when the values of the dependent variable are restricted to a range of values. As in the case of truncation the dependent variable is only observed for a subsample. However, there is information (the independent variables) about the whole sample.

Some examples:

- Income data are often *top-coded* in survey data. For example, all incomes above CHF 200,000 may be reported as CHF 200,000. However, households with high incomes are part of the sample and their characteristics reported.
- Tickets sold for soccer matches cannot exceed the stadion's capacity.
- Expenditures for durable goods are either positive or zero. (This is the example used in Tobin's (1958) original paper.)
- The number of extramarital affairs are nonnegative. (Note that although Fair's (1978) famous article uses a Tobit model, count data models may be more appropriate)

2.1 The Model (Tobit Type 1)

Consider a latent random variable y_i that linearly depends on x_i, i.e.

$$y_i^* = x_i'\beta + \varepsilon_i \quad \text{with} \quad \varepsilon_i|x_i \sim N(0, \sigma^2).$$

The error term ε_i is independently and normally distributed with mean 0 and variance σ^2. The distribution of y_i given x_i is therefore also normal: $y_i^*|x_i \sim N(x_i'\beta, \sigma^2)$. The expected value of the latent variable is $E(y_i^*|x_i) = x_i'\beta$.

The observed value y_i is censored below 0, i.e.[2]

$$y_i = \begin{cases} y_i^* & \text{if } \; y_i^* > 0 \\ 0 & \text{if } \; y_i^* \leq 0 \end{cases}$$

The observed variable is a mixture random variable with a probability mass $P(y_i = 0|x_i) = P(y_i^* < 0|x_i) = \Phi(-x_i'\beta/\sigma)$ on 0 and a continuum of values above 0 with density $f(y_i|x_i) = \sigma\phi[(y_i - x_i'\beta)/\sigma]$.

The expected value of the observed variable is

$$\begin{aligned} E(y_i|x_i) &= 0 \cdot P(y_i^* \leq 0|x_i) + E(y_i^*|y_i^* > 0, x_i) \cdot P(y_i^* > 0|x_i) \\ &= \left[x_i'\beta + \sigma\frac{\phi\left(x_i'\beta/\sigma\right)}{\Phi\left(x_i'\beta/\sigma\right)}\right]\Phi\left(x_i'\beta/\sigma\right) \\ &= x_i'\beta\,\Phi\left(x_i'\beta/\sigma\right) + \sigma\phi\left(x_i'\beta/\sigma\right) \end{aligned}$$

Figure 2 visualizes the truncated regression model in an example.

2.2 Interpretation of Parameters

The interpretation of the parameters depends very much on research question. If the researcher is interested in the underlying linear relationship of the whole population, the slope coefficients β can simply be interpreted as marginal effects

$$\frac{\partial E\left(y_i^*|x_i\right)}{\partial x_{ik}} = \beta_k$$

However, if the researcher is interested in the effect on the expected value of the observed (censored) value, the marginal effect is (derive!)

$$\frac{\partial E\left(y_i|x_i\right)}{\partial x_{ik}} = \beta_k\,\Phi\left(\frac{x_i'\beta}{\sigma}\right)$$

[2] It is straightforward to study any known treshold $a \neq 0$ within the above framework. If the original variable y_i is censored below at a then $z_i = y_i - a$ satisfies the Tobit model. If the original variable is censored above at a, then $z_i = -(y_i - a)$ is a standard Tobit model.

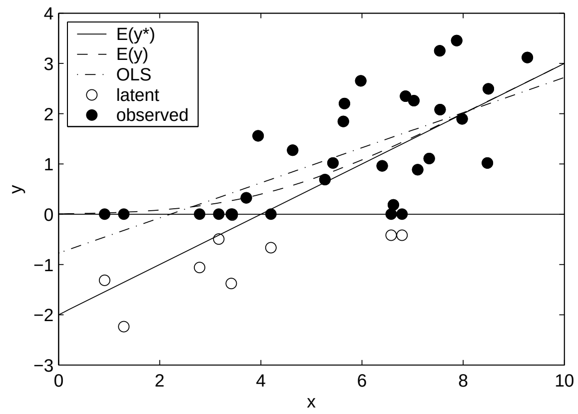

Figure 2: The standard (type 1) Tobit model. Lower censoring at 0, $N = 30$, $K = 2$, $\beta = (-2, 0.5)'$ and $\sigma = 1$

There is an interesting decomposition of this marginal effect (McDonald and Mofitt, 1980): (1) the effect on the expectation of fully observed values and (2) the effect on the probability of being fully observed:

$$\frac{\partial E\left(y_i|x_i\right)}{\partial x_{ik}} = \underbrace{\frac{\partial E\left(y_i^*\,|y_i^* > 0, x_i\right)}{\partial x_{ik}}}_{(1)} P(y_i^* > 0) + \underbrace{\frac{\partial P(y_i^* > 0)}{\partial x_{ik}}}_{(2)} E\left(y_i^*\,|y_i^* > 0, x_i\right)$$

with

$$\frac{\partial E\left(y_i^*\,|y_i^* > 0, x_i\right)}{\partial x_{ik}} = \beta_k\left(1 - \lambda_i^2 - \alpha_i\lambda_i\right)$$

$$\frac{\partial P(y_i^* > 0)}{\partial x_{ik}} = \frac{\partial\Phi(x_i'\beta/\sigma)}{\partial x_{ik}} = \beta_k\,\sigma^{-1}\phi(x_i'\beta/\sigma)$$

where $\lambda_i \equiv \phi(-x_i'\beta/\sigma)/[1 - \Phi(-x_i'\beta/\sigma)] = \phi(x_i'\beta/\sigma)/\Phi(x_i'\beta/\sigma)$ and $\alpha_i = x_i'\beta/\sigma$. These marginal effects depend on individual characteristics x_i and can only be reported for specified types or as average effects in the sample population.

2.3 Estimation

The OLS regression of the observed variable y_i on x_i

$$y_i = x_i'\beta + u_i$$

will yield biased estimates of β, as $E(y_i|x_i) = x_i'\beta\,\Phi(\alpha_i) + \sigma\phi(\alpha_i)$ is not a linear function of x_i. Note that restricting the sample to fully observed observations, i.e. where $y_i > 0$, does not solve the problem as can be seen in the truncated regression model above.

The censored regression is usually estimated by maximum likelihood (ML). Assuming independence across observations, the log likelihood function is

$$\ln \mathcal{L} = \sum_{\{i|y_i>0\}} \ln\left[\sigma^{-1}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right] + \sum_{\{i|y_i=0\}} \ln\left[1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)\right]$$

and allows to estimate both β and σ by an iterative numerical procedure. The above likelihood function is a (strange) mixture of discrete and continuous components and standard ML proofs do not apply. However, it can be shown that the Tobit estimator has the usual ML properties: consistency, asymptotic efficiency and normality and asymptotic hypothesis tests can be performed as Wald, likelihood ratio or lagrange multiplier tests. Although the log-likelihood function of the Tobit model is not globally concave, it has a unique maximum.

The ML estimation of the of the censored regression model rests on the strong assumption that the latent error term is normally distributed and homoscedastic. The ML estimator is inconsistent in the presence of heteroscedasticity and robust covariance estimators cannot solve this. Several *semi-parametric* estimation strategies have been proposed that relax the distributional assumption about the error term. See Chay and Powell (2001) for an introduction.

2.4 Implementation in STATA 10.0

Stata estimates the standard (type 1) tobit model by the command

```
tobit depvar [indepvars], ll(0)
```

More general models with censoring from above and below are estimated by

```
tobit depvar [indepvars], ll(#) ul(#)
```

where the number in `ll(#)` in the lower threshold and then number in `ul(#)` is the upper threshold.

The post-estimation commands `predict` calculates by default the linear index function $x_i'\beta$. The option `predict, ystar(0,.)` predicts the observed censored variable $E(y_i|x_i) = E(y_i^*|y_i^* > 0, x_i) \cdot P(y_i^* > 0|x_i)$. The post-estimation command `mfx` reports marginal effects. By default $\partial E(y_i^*|x_i)/\partial x_{ik} = \beta_k$ is reported. The marginal effect on the observed censored variable, $\partial E(y_i|x_i)/\partial x_{ik}$, is calculated by `mfx, predict(ystar(0,.))`. See `help tobit_postestimation` on how to calculate all parts of the McDonald and Mofitt decomposition using `predict` and `mfx`.

For example,

```
tobit housing inc age edu, ll(0)
predict housing_hat, ystar(0,.)
mfx, predict(ystar(0,.)) at(inc=50000, age=45, edu=12)
```

predicts $E(y_i|x_i)$ and calculates the marginal effects of income, age and experience on expected observed housing expenditures $E(y_i|x_i)$ for a 45 year old person with income 50,000 and 12 years of education.

3 Selection

The sample selection problem occurs when the observed sample is not a random sample from the whole population but from a distinct subset of the population. Truncation and censoring as special cases are special cases of sample selection or *incidental truncation*.

The classical example: Income is only observed for employed persons but not for the ones that decide to stay at home (historically mainly women).

3.1 The Model (Heckman Selection Model, Tobit Type 2)

Consider a model with two latent variables y_i^* and d_i^* which linearly depend on observable independent variables x_i and z_i, respectively

$$\begin{aligned} d_i^* &= z_i'\gamma + \nu_i \\ y_i^* &= x_i'\beta + \varepsilon_i \end{aligned}$$

with

$$\nu_i, \varepsilon_i | x_i, z_i \sim N\left(0, \begin{bmatrix} 1 & \rho\,\sigma_e \\ \rho\,\sigma_e & \sigma_\varepsilon^2 \end{bmatrix}\right)$$

The error terms ε_i and ν_i are independently (across observations) and jointly normally distributed with covariance $\rho\,\sigma_\varepsilon$. Note that the variance of ν_i is set to unity as it is *not identified* in the estimation.

The two latent variables cannot be observed by the researcher. She only observes an indicator d_i when the latent variable d_i^* is positive. The value of the variable $y_i = y_i^*$ is only observed if the indicator is 1:

$$\begin{aligned} d_i &= \begin{cases} 1 & \text{if } d_i^* > 0 \\ 0 & \text{otherwise} \end{cases} \\ y_i &= \begin{cases} y_i^* & \text{if } d_i = 1 \\ \text{n.a.} & \text{otherwise} \end{cases} \end{aligned}$$

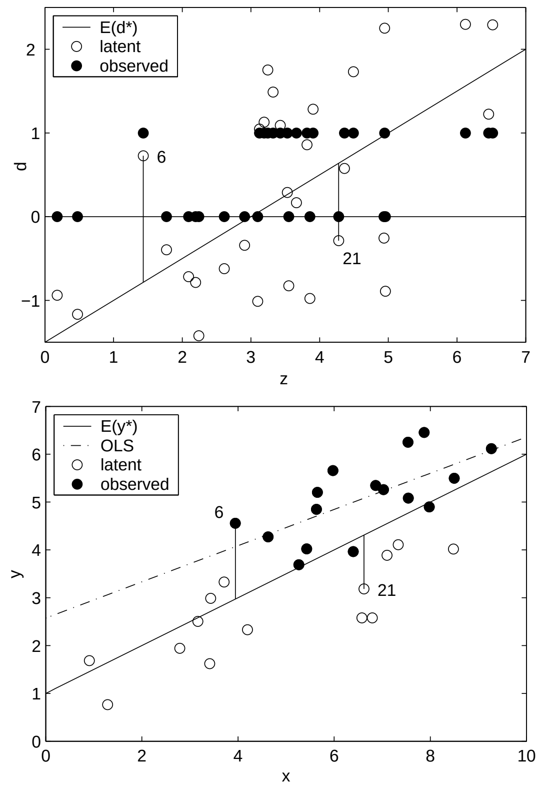

Figure 3: The selection model with correlated observable and unobservable characteristics. $N = 30$, $\gamma = (-1.5, 1)'$ $\beta = (-2, 0.5)'$, $\sigma_\varepsilon = 1$, $\rho = 0.8$ and $corr(x_i, z_i) = 0.5$.

In other words, the first equation (the decision equation d_i^*) explains whether an observation is in the sample or not. The second equation (the regression equation y_i^*) determines the value of y_i. Note that the standard tobit model is a special case of this setup with $z_i = x_i$, $\gamma = \beta$, $\sigma_\nu = \sigma_\varepsilon$ and $\rho = 1$.

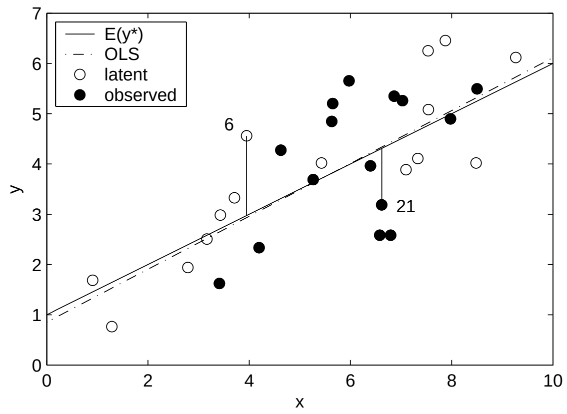

Figure 4: The selection model with correlated observable but *un*correlated unobservable characteristics. $\rho = 0$.

Figure 3 shows an example of a selection model. The positive correlation between x and z explains why the probability of being observed increases with x. The positive error correlation explains why, for given x_i and z_i, points y_i^* above the expected value (e.g. point 6) are more likely to be observed.

The expected value of the variable y_i is the conditional expectation of y_i^* conditioned on it being observed ($d_i = 1$)

$$E(y_i|x_i, z_i) = E(y_i^*|d_i = 1, x_i, z_i) = x_i'\beta + \rho\,\sigma_\varepsilon \frac{\phi(z_i'\gamma)}{\Phi(z_i'\gamma)} = x_i'\beta + \rho\,\sigma_\varepsilon \lambda(z_i'\gamma)$$

where $\lambda(\alpha) \equiv \phi(\alpha)/\Phi(\alpha)$ is called the *inverse Mills ratio*.

Note that $E(y_i|x_i, z_i) = x_i'\beta$ if the two error terms are uncorrelated, i.e. $\rho = 0$. This is yet true when x_i and z_i are correlated, as for example in the usual case when some independent variables appear in x and z.

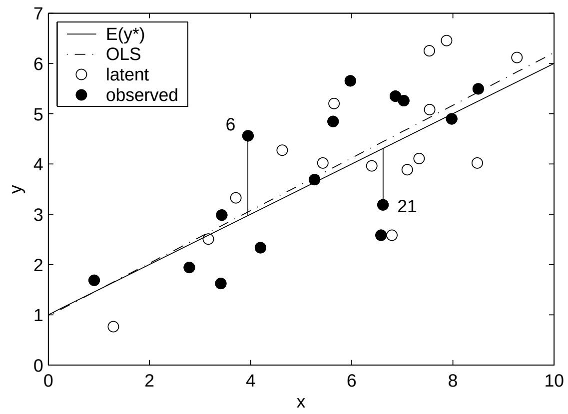

Figure 5: The selection model with both uncorrelated observable and uncorrelated unobservable characteristics, i.e random sampling. $\rho = 0$ and $corr(x_i, z_i) = 0$.

3.2 Interpretation of Parameters

In most cases, we are interested on the effect of independent variables in the whole population. Therefore we would like to obtain an unbiased and consistent estimator of β which is directly interpreted as marginal effect.

In some cases, however, the researcher is interested in the effect on the observed population. For regressors that appear on the LHS of both y_i^* and d_i^*, the marginal effect depends not only on β but also on γ through the probability of being in the sample. See your textbook for details.

3.3 Estimation

OLS regression of the observed variable y_i on x_i

$$y_i = x_i'\beta + u_i$$

will yield biased estimates of β as the factor $\rho\sigma_\varepsilon\phi(z_i'\gamma)/\Phi(z_i'\gamma)$ is omitted and becomes part of the error term. The error term u_i is therefore correlated with x_i if $\rho \neq 0$ and z_i is correlated with x_i. The resulting bias is called *selection bias* or *sample selectivity bias*.

Note that there is no bias if the unobservable components are uncorrelated ($\rho = 0$) even when the observed sample is highly selective, i.e. even when x and z are correlated and thus some values of x are more likely to be observed than others. Figure 4 shows this situation. Needless to say that there is no bias if the observable and unobservable characteristics between the decision and the regression equation are uncorrelated. This case of a pure random sample is sketched in Figure 5.

3.3.1 Estimation with Maximum Likelihood

The decision and regression equations can be simultaneously estimated by maximum likelihood under the distributional assumptions made. The log-likelihood function consists of two parts: (1) The likelihood contribution from observations with $d_i = 0$, i.e. the probability of not being observed in the regression equation. (2) The likelihood contribution from observations with $d_i = 1$, i.e. the probability of being observed multiplied with the conditional density of the observed value.

$$\begin{aligned}
\ln\mathcal{L} &= \sum_{d_i=0} \ln P(d_i = 0) + \sum_{d_i=1} \ln\left[P(d_i = 1) f(y_i^*|d_i = 1)\right] \\
&= \sum_{d_i=0} \ln P(d_i = 0) + \sum_{d_i=1} \ln\left[f(y_i^*) P(d_i = 1|y_i^*)\right] \\
&= \sum_{d_i=0} \ln P(d_i = 0) + \sum_{d_i=1} \ln f(y_i^*) + \sum_{d_i=1} \ln P(d_i = 1|y_i^*) \\
&= \sum_{d_i=0} \ln\left[\Phi(-z_i'\gamma)\right] + \sum_{d_i=1} \ln\left[\sigma^{-1}\phi\left((y_i - x_i'\beta)/\sigma\right)\right] \\
&\quad + \sum_{d_i=1} \ln\left[\Phi\left(\frac{z_i'\gamma + \rho\sigma^{-1}(y_i - x_i'\beta)}{(1-\rho^2)^{1/2}}\right)\right]
\end{aligned}$$

Note that this likelihood function identifies β, γ, ρ, σ_ε but not the variance of ν which was set to unity. Although β and γ are theoretically identified they are difficult to identify in practice when the same variables are included in both equations, i.e. when $x_i = z_i$. It is therefore strongly advised to include variables in z that are not included in x. In the case of $\rho = 0$, the log likelihood functions reduces to the sum of a probit and a standard linear regression model which can be estimated separately.

The ML estimation of the selection model has standard ML properties (consistency, efficiency, asymptotic normality, etc). In practice it is often difficult to numerically find the maximum values and good starting values are very important. Therefore, estimates from the two-step procedure in the following section are often used as starting values. The ML estimation is only necessary when a test on $\rho = 0$ is rejected in the two-step estimation.

The ML estimation of the heckman selection model rests heavily on the assumption that the error terms are *jointly* normally distributed. This is a very strong and often unrealistic assumption. Several *semi-parametric* estimation strategies have been proposed that relax the distributional assumption about the error term. See Vella (1998) for an introduction.

3.3.2 Estimation with Heckman's Two-Step Procedure

Heckman proposed a two-step procedure which only involves the estimation of a standard probit and a linear regression model. The two step procedure draws on the conditional mean

$$E(y_i|x_i, z_i) = x_i'\beta + \rho\,\sigma_\varepsilon \frac{\phi(z_i'\gamma)}{\Phi(z_i'\gamma)} = x_i'\beta + \rho\,\sigma_\varepsilon \lambda(z_i'\gamma)$$

of the fully observed y's.

Step 1 is the consistent estimation of γ by ML using the full set of

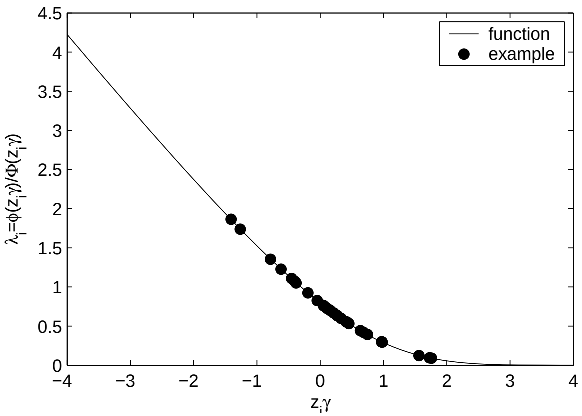

Figure 6: The inverse Mills ratio and observations from Figure 3.

observations in the standard probit model

$$\begin{aligned} d_i^* &= z_i'\gamma + \nu_i \\ d_i &= 1 \text{ if } \mathrm{d}_\mathrm{i}^* > 0, 0 \text{ otherwise.} \end{aligned}$$

We can use this to consistently estimate the inverse Mills ratio $\hat{\lambda}_i = \phi(z_i'\hat{\gamma})/\Phi(z_i'\hat{\gamma})$ for all observations.

Step 2 is the estimation of the regression equation with the inverse Mills ratio as an additional variable

$$y_i = x_i'\beta + \beta_\lambda \hat{\lambda}_i + u_i$$

for the subsample of full observations. The OLS regression yields $\hat{\beta}$, $\hat{\beta}_\lambda$, $\hat{\sigma}_\varepsilon$ and thus the correlation $\hat{\rho} = \hat{\beta}_\lambda/\hat{\sigma}_\varepsilon$.

Heckman's two step estimator is consistent but not efficient. Furthermore, the covariance matrix of the second-step estimator provided by standard OLS is incorrect as one regressor (the Mills ratio) is measured with error and the error term u_i is heteroskedastic. Therefore the

standard errors need to be corrected. However, the test on the null hypothesis $\beta_\lambda = 0$ whic is a test on $\rho = 0$ can be performed using the "incorrect" OLS standard errors (as they are "correct" under the null hypothesis).

There is often a practical problem of identification (almost multicollinearity) when the variables in both equations are the same, i.e. $x_i = z_i$ (See Vella, 1998). The parameters β and β_λ are theoretically identified by the non-linearity of the inverse Mills ratio $\lambda(.)$. However, as can be seen in Figure 6, $\lambda(.)$ is almost linear for a large range of values $z_i'\gamma$. It is therefore strongly advised to include variables in z that are not included in x.

3.4 Implementation in STATA 10.0

Stata calculates ML estimates the by the command

```
heckman depvar varlist, select(depvar_s = varlist_s)
```

where in our notation `depvar` $= y$, `varlist` $= x$, `depvar_s` $= d$ and `varlist_s` $= z$. Stata calculates two-step estimates with corrected standard errors by adding the option `twostep`.

References

Cameron, A. C. and P. K. Trivedi (2005), *Microeconometrics: Methods and Applications*, Cambridge University Press. Chapter 16.

Wooldridge, Jeffrey M. (2009), *Introductory Econometrics: A Modern Approach*, 4th ed. South-Western. Chapter 17.4 and 17.5.

Chay, Kenneth Y. and James L. Powell (2001), Semiparametric Censored Regression Models, *Journal of Economic Perspectives*, 15(4), 29-42.

Vella, F. (1998) Estimating Models with Sample Selection Bias: A Survey, *Journal of Human Resources*, 33, 127-169.