1 The Type I Tobit Model

$$
\begin{aligned}
y_i^* &= \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i, \\
y_i &= y_i^* \text{ if } y_i^* > 0 \\
y_i &= 0 \text{ if } y_i^* \leqslant 0
\end{aligned}
$$

Wooldridge distinguishes between two different applications of this model:

1. Censored data.

"Top coding" typical example:

$$
\begin{aligned}
E\left(wealth^* \mid \mathbf{x}\right) &= \mathbf{x}' \boldsymbol{\beta} \\
wealth &= \min\left(wealth^*, 200\right) \\
wealth^* &= \mathbf{x}' \boldsymbol{\beta} + u,\ u \mid \mathbf{x} \sim \mathbf{N}\left(0, \sigma^2\right)
\end{aligned}
$$

2. Corner solution outcomes

Example: Engel curves for alcohol and tobacco consumption:

$$
w_j = \alpha_{ji} + \beta_{ji} \log x_i + \varepsilon_{ji},
$$

where w_j is the budget share of commodity j, β_{ji} is the income elasticity that can depend on household characteristics, x is income.

$$
\begin{aligned}
w_{ji}^* &= \alpha_{ji} + \beta_{ji} \log x_i + \varepsilon_{ji}, \\
w_{ji} &= w_{ji}^* \text{ if } w_{ji}^* > 0 \\
&= 0 \text{ otherwise.}
\end{aligned}
$$

Different outcomes from this model can be interesting, e.g.:

1. $P(y=0)$ and $P(y>0)$
2. $E(y \mid \mathbf{x})$
3. $E(y \mid \mathbf{x}, y>0)$

$P(y=0)$ is trivial. Remember from the threshold model

$$\begin{aligned} P(y_i = 0) &= P(y_i^* \leqslant 0) = P(\varepsilon_i \leqslant -\mathbf{x}_i'\boldsymbol{\beta}) = \\ P\left(\frac{\varepsilon_i}{\sigma} \leqslant \frac{\mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right) &= \Phi\left(-\frac{\mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right) = 1 - \Phi\left(\frac{\mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right). \end{aligned}$$

Implies that

$$P(y_i > 0) = \Phi\left(\frac{\mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right).$$

This allows for that we can estimate $\boldsymbol{\beta}$ and σ consistently, but not separately.

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

To derive $E(y \mid \mathbf{x}, y>0)$ we need a result from the theory of truncated normal distributions.

1.1 Truncated normal distributions

Censored random variables:

$$y_i = y_i^* \text{ if } y_i^* > c$$
$$y_i = c \text{ if } y_i^* \leqslant c$$

Truncated random variables:

$$y_i = y_i^* \text{ if } y_i^* > c$$
$$\text{nothing if } y_i^* \leqslant c$$

Truncated normal

If the distribution is truncated from below.

$$\begin{aligned} f(y|Y \geqslant c) &= \frac{f(y)}{P(Y \geqslant c)} \text{ if } y \geqslant c \\ & \quad 0 \text{ otherwise} \end{aligned}$$

If $Y \sim N(\mu, \sigma^2)$ then

$$E(Y|Y \geqslant c) = \mu + \sigma\lambda_1(c^*) \geqslant \mu,$$

where $\lambda_1(c) = \frac{\phi(c)}{1-\Phi(c)}$ and $c^* = \frac{c-\mu}{\sigma}$.

The corresponding result when the distribution is truncated from above:

$$E(Y|Y \geqslant c) = \mu + \sigma\lambda_2(c^*) \leqslant \mu,$$

where $\lambda_2(c) = \frac{-\phi(c)}{\Phi(c)}$.

For a bivariate distribution (Y, X) :

$$\begin{aligned} E(Y|X \geqslant c) &= \mu_y + \left(\sigma_{yx}/\sigma_x^2\right)[E(X|X \geqslant c) - \mu_x] \\ &= \mu_y + (\sigma_{yx}/\sigma_x)\lambda_1(c^*) \end{aligned}$$

Applying this result to the Tobit model:

$$E\left(y_i|\mathbf{x}_i, y_i > 0\right) = \mathbf{x}_i'\boldsymbol{\beta} + E\left(\varepsilon_i|\varepsilon_i > -\mathbf{x}_i'\boldsymbol{\beta}\right) = \mathbf{x}_i'\boldsymbol{\beta} + \sigma\frac{\phi\left(-\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right)}{1 - \Phi\left(-\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right)} = \mathbf{x}_i'\boldsymbol{\beta} + \sigma\frac{\phi\left(\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right)}{\Phi\left(\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right)}.$$

For any point c, $\frac{\phi(c)}{\Phi(c)}$ is called the inverse of Mill's ratio $\lambda\left(c\right)$.

Marginal effect:
It can be shown that

$$\frac{d\lambda\left(c\right)}{dc} = -\lambda\left(c\right)\left[c + \lambda\left(c\right)\right]$$

$$\frac{\partial E\left(y_i \mid \mathbf{x}_i, y_i > 0\right)}{\partial x_j} = \beta_j\left[1 - \lambda\left(\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right)\left[\mathbf{x}_i'\boldsymbol{\beta}/\sigma + \lambda\left(\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right)\right]\right],$$

where the adjustment factor $\left[1 - \lambda\left(\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right)\left[\mathbf{x}_i'\boldsymbol{\beta}/\sigma + \lambda\left(\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right)\right]\right]$ is always between 0 and 1. This means that β_j determines the sign of the partial effect.

$$\begin{aligned} E\left(y_i|\mathbf{x}_i\right) &= E\left(y_i|\mathbf{x}_i, y_i > 0\right) P\left(y_i > 0\right) + 0 * P\left(y_i = 0\right) = \\ &= \left[\mathbf{x}_i'\boldsymbol{\beta} + \sigma\frac{\phi\left(\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right)}{\Phi\left(\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right)}\right]\Phi\left(\frac{\mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right) = \\ &= \mathbf{x}_i'\boldsymbol{\beta}\Phi\left(\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right) + \sigma\phi\left(\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right) \end{aligned}$$

Decomposition by McDonald and Moffitt (1980) of partial effect:

$$\frac{\partial E\left(y_i|\mathbf{x}_i\right)}{\partial x_j} = \frac{\partial P\left(y_i > 0 \mid \mathbf{x}_i\right)}{\partial x_j} E\left(y_i|\mathbf{x}_i, y_i > 0\right) + P\left(y_i > 0 \mid \mathbf{x}_i\right)\frac{\partial E\left(y_i|\mathbf{x}_i, y_i > 0\right)}{\partial x_j}.$$

First component shows the effect on the observed segment and the second component on probability of exceeding the threshold.

This can be simplified to the following expression:

$$\frac{\partial E\left(y_i \mid \mathbf{x}_i\right)}{\partial x_{ij}} = \beta_j\Phi\left(\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right)$$

The second component, $\Phi\left(\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right)$, is the probability of observing the dependent variable. So the adjustment is larger if there is a large probability of observing is small, which is intuitive.

Note that any ratio between coefficients, such as β_j/β_k, are, as in the logit/probit case,equal to the ratios between the marginal effects.

Inconsistency of OLS:

The source of the inconsistency can be seen as a specification error. Suppose we only include the non-zero observations, Wooldridge write the model as:

$$\begin{aligned} y_i &= \mathbf{x}_i'\boldsymbol{\beta} + \boldsymbol{\sigma\lambda}\left(\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right) + e_i \\ E\left(e_i \mid \mathbf{x}_i, y_i > 0\right) &= 0 \end{aligned}$$

Omitted variable bias will occur if there is a correlation between $\boldsymbol{\lambda}\left(\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right)$ and $\mathbf{x}_i$.

Wooldridge also points out that it is unlikely that running OLS on all data will give consistent estimates, since $E\left(y_i \mid \mathbf{x}_i\right)$ is non-linear in $\mathbf{x}_i$ and $\boldsymbol{\beta}$.

1.2 Estimation

Estimated by maximum likelihood

$$L = \prod_{i \in I_0} P(y_i = 0) \prod_{i \in I_1} f(y_i | y_i > 0) P(y_i > 0)$$

$$\begin{aligned}
\ln L &= \sum_{i \in I_0} \ln P(y_i = 0) + \sum_{i \in I_1} [\ln f(y_i | y_i > 0)] + \ln P(y_i > 0) \\
&= \sum_{i \in I_0} \ln P(y_i = 0) + \sum_{i \in I_1} \ln f(y_i) = \\
&= \sum_{i \in I_0} \ln [1 - \Phi((\mathbf{x}_i' \boldsymbol{\beta} / \sigma))] + \sum_{i \in I_1} \ln \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-(y_i - \mathbf{x}_i' \boldsymbol{\beta})^2 / 2\sigma^2 \right) \right]
\end{aligned}$$

1.3 Specification and test for heteroscedasticity

Note that the Tobit model is fairly restrictive specification. It is required that the independent variables should have the same effect on the probability of being observed in the sample as in the observed segment of the model. As in the logit/probit case, heteroscedasticity will lead to inconsistent estimates of the $\boldsymbol{\beta}$ vector. Lots of Monte Carlo studies on that. Wooldridge discusses a LM test for heteroscedasticity.

An informal mis-specification test is the estimate the probit part separately. This gives an estimate of $\boldsymbol{\beta}/\sigma$, which can be compared to the separate estimates of $\boldsymbol{\beta}$ and σ from the Tobit model. If they are statistically different we can conclude a mis-specification. Like a pseudo Hausman test.

Heteroscedasticity can be very problematic in Tobit models.

Results from Mote Carlo study in Johnston nd DiNardo:
Model:

$$y_i^* = x_i - 10 + \epsilon_i,$$

where $\epsilon_i \sim N\left(0, x_i^2\right)$. 200 replicates equally spaced from 0.02 to 40.
Results:

	Mean	SD	Min	Max
Percent censored	.423	.029	.340	.495
OLS slope	.975	.125	.631	1.427
Tobit slope	1.678	.211	1.133	2.329

LM test for heteroscedasticity (similar to that for probit models):

$$Var\left(u \mid \mathbf{x}\right) = \sigma^2 \exp\left(\mathbf{z}\boldsymbol{\delta}\right),$$

where $\mathbf{z}$ is a subvector of $\mathbf{x}$. Can be estimated under $H_0 : \boldsymbol{\delta} = \mathbf{0}$.

1.4 Alternatives to the Tobit model

1.4.1 Parametric Alternatives

Alternative to the Tobit model, which relaxes the assumption that it is the same index function determining censoring and levels, is the "hurdle model" or "two-tired model". Discussed in Wooldridge in Section 16.7. Wooldridge suggests a model with a probit equation determining participation and a log-normal distribution above the censoring point, i.e.,

$$P(y=0 \mid \mathbf{x}) = 1 - \Phi(\mathbf{x}\boldsymbol{\gamma})$$

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

From that we obtain:

$$\begin{aligned} f(y \mid \mathbf{x}) &= P(w=0 \mid \mathbf{x}) f(y \mid \mathbf{x}, w=0) + P(w=1 \mid \mathbf{x}) f(y \mid \mathbf{x}, w=1) = \\ &= 1[w=0][1-\boldsymbol{\Phi}(\mathbf{x}\boldsymbol{\gamma})] + 1[w=1]\boldsymbol{\Phi}(\mathbf{x}\boldsymbol{\gamma})\phi[\{\log(y) - \mathbf{x}\boldsymbol{\beta}\}/\sigma]/(y\sigma), \end{aligned}$$

where w is an indicator for $y > 0$.

The maximum likelihood function is obtained in a parallel fashion as in the Tobit I case.

Cragg (1971) suggested a similar model where the log-normal part of the model above is replaced by a truncated normal.

1.4.2 Semi-parametric Alternatives

Symetrically Trimmed Least squares Powell (1984). Idea: Censor the observations in the right tail as well.

$$y_i = \begin{cases} y_i^* \text{ if } y_i^* > 0 \\ 0 \text{ if } y_i^* \leqq 0 \end{cases}$$

or equivalently:

$$y_i = \begin{cases} y_i^* \text{ if } \epsilon_i > -\mathbf{x}_i'\boldsymbol{\beta} \\ 0 \text{ if } \epsilon_i \leqq -\mathbf{x}_i'\boldsymbol{\beta} \end{cases}$$

Censoring a problem since it introduces an asymmetry. Iterative process:

1. Compute an initial estimate of $\boldsymbol{\beta}$ with OLS.

2. Compute the predicted value

   - If the predicted value is negavive, set the observation to 0.
   - If the value of the dependent variable is twice as big as the predicted value, set the dependent variable to $2\mathbf{x}_i'\boldsymbol{\beta}$.

3. Run OLS on altered data.

4. Repeat until $\boldsymbol{\beta}$ stop changing.

Censored Least Absolute Deviations (CLAD) Estimator Suggested by Powell (1984). Instead of the least squared deviations, use the least absolute ones.

$$\min_{\widehat{\beta}} \left[\sum |y_i^* - \mathbf{x}_i'\boldsymbol{\beta}| \right] = \min_{\widehat{\beta}} \sum (y_i^* - \mathbf{x}_i'\boldsymbol{\beta}) \cdot sign\,(y_i^* - \mathbf{x}_i'\boldsymbol{\beta})\,.$$

With the normal equation:

$$\mathbf{0} = \sum \mathbf{x}_i' \cdot sign\,(y_i^* - \mathbf{x}_i'\boldsymbol{\beta})\,.$$

This is only dependent on the sign of the predicted residual and it can be shown that it corresponds to the median regression, i.e.,

$$q_{50}\left[y_i^* \mid \mathbf{x}_i\right] = \mathbf{x}_i'\boldsymbol{\beta} + q_{50}\left[\epsilon_i \mid \mathbf{x}_i\right] = \mathbf{x}_i'\boldsymbol{\beta}.$$

2 Selective samples and the Tobit II model

2.1 When is selection no problem?

Section 17.2 in Wooldridge.

Model:

$$y = \beta_1 + \beta_2 x_2 + ... + \beta_K x_K + u.$$

An instrument $\mathbf{z}$ is used such that $E(u \mid \mathbf{z}) = 0$.

s is a selection indicator such that $s = 1$ if the individual is observed and $s = 0$ otherwise.

Key assumption: $E(u \mid \mathbf{z}, s) = 0$, which implies that $E(s_i \mathbf{z}_i u_i) = 0$.

Theorem 17.1 states that OLS gives consistent estimates under this + the usual 2SLS assumptions.

Example from famous paper by Griliches, Hall and Hausman.

$$\log(wage) = \mathbf{z}_1 \delta_1 + abil + v,$$

where $abil = \theta_1 IQ + e$, where $E(e \mid \mathbf{z}_1, IQ) = 0$ and $E(v \mid \mathbf{z}_1, abil, IQ) = 0$.

Thus,

$$\log(wage) = \mathbf{z}_1 \delta_1 + \theta_1 IQ + u.$$

Probability of missing is higher at lower IQ levels.

$$s = 1 \text{ if } IQ + r \geq 0; \;\; s = 0 \text{ if } IQ + r < 0,$$

where r is an unobserved random variable. If $E(v \mid \mathbf{z}_1, abil, IQ, r) = 0$, $E(u \mid \mathbf{z}_1, abil, IQ, r) = 0$ and, since s is a function of r and IQ, $E(u \mid \mathbf{z}_1, IQ, s) = 0$, the model can be estimated consistently on the observed sample.

This would, however, not apply if say r and e were related, i.e., if say those with e.g. low ability were more prone to drop out.

2.2 Truncation on the basis of y

Wooldridge section 17.3.

Sometimes it is only possible to observe the dependent variable in a particular interval, e.g. top coding or evaluations of programs available to low income earners. The latter example applies to Hausman and Wise's study of negative income tax.

Difference from censored models: We cannot observe $\mathbf{x}_i$ for those not included in the sample.

Underlying model:

$$E(y_i \mid \mathbf{x}_i) = \mathbf{x}_i \boldsymbol{\beta}$$

$$s_i = 1[a_1 < y_i < a_2]$$

a_1 and a_2 are known constants. It is straight forward to obtain the probability density function for the truncated variable:

$$p(y \mid \mathbf{x}_i, s_i = 1) = \frac{f(y \mid \mathbf{x}_i; \boldsymbol{\beta})}{F(a_2 \mid \mathbf{x}_i; \boldsymbol{\beta}) - F(a_1 \mid \mathbf{x}_i; \boldsymbol{\beta})} = \frac{f(\mathbf{x}_i' \boldsymbol{\beta})}{F(a_2 - \mathbf{x}_i' \boldsymbol{\beta}) - F(a_1 - \mathbf{x}_i' \boldsymbol{\beta})}$$

The log-likelihood function for the model could then be obtained from this expression.

2.3 Bounds

Manski (1995) "Identification Problems in the Social Sciences"

y is an outcome, x a conditioning variable ans z a selection indicator (=1 if the individual is observed).

Bounds with no prior information:

$$\begin{aligned} P(y &= 1 \mid x, z=1)P(z=1 \mid x) \leqslant \\ P\left(y=1 \mid x\right) &\leqslant \\ P(y &= 1 \mid x, z=1)P(z=1 \mid x) + P(z=0 \mid x) \end{aligned}$$

Bounds with ordered outcomes ($y_1 \geqslant y_0$):

$$\begin{aligned} P(y &= 1 \mid x, z=1)P(z=1 \mid x) \leqslant \\ P\left(y=1 \mid x\right) &\leqslant \\ P(y &= 1 \mid x, z=1)P(z=1 \mid x) + P(z=0 \mid x)P(y=1 \mid x, z=1) \end{aligned}$$

Treatment with larger outcome.
For any cutoff poit t

$$\begin{aligned} P(y_1 &\leq t \mid x, z=0) \geq P(y_0 \leq t \mid x, z=0) \text{ then} \\ P\left(y=1 \mid x\right) &= P(y_1 \leq t \mid x, z=1)P(z=1 \mid x) + P(y_1 \leq t \mid x, z=0)P(z=0 \mid x) \\ &\geq P(y_1 \leq t \mid x, z=1)P(z=1 \mid x) + P(y_0 \leq t \mid x, z=0)P(z=0 \mid x) \\ &= P(y \leq t \mid x) \end{aligned}$$

Blundell, Gosling, Ichimura and Meghir "Changes in the distribution of Male and Female Wages Accounting for Employment Composition Using Bounds" *Econometrica* 2007.

Analyses the income distribution.

Bounds with no prior information:

$$F(w \mid x) = F(w \mid x, E=1)P(x) + F(w \mid x, E=0)\left[1 - P(x)\right]$$

Use the fact that $0 \leq F(w \mid x, E=0) \leq 1$

$$F(w \mid x, E=1)P(x) \leq F(w \mid x) \leq F(w \mid x, E=1)P(x) + \left[1 - P(x)\right]$$

This gives us

$$w^{q(l)}\left(x\right) \leq w^{q}\left(x\right) \leq w^{q(u)}\left(x\right)$$

Stochastic dominance of those who are employed. The earnings distribution of those who work stochastically dominates those who do not work, i.e.,

$$F(w \mid x, E=1) \leq F(w \mid x, E=0)$$

or equivalently

$$\Pr\left(E=1 \mid W \leq w, x\right) \leq \Pr\left(E=1 \mid W > w, x\right),$$

i.e., you will find more employed people higher upp in the earnings distribution. This restriction implies that the bounds can be redefined as

$$F\left(w \mid x, E=1\right) \leq F(w \mid x) \leq F\left(w \mid x, E=1\right)P(x) + \left[1 - P(x)\right].$$

Median restriction. The wage offers of those not working are not higher than the median earning of those who do work, i.e.,

$$\begin{aligned} 0 &\leq F\left(w \mid E=0\right) \leq 1 \qquad \text{if } w < w^{50(E=1)}(x) \\ 0.5 &\leq F(w \mid x, E=0) \leq 1, \quad \text{if } w \geq w^{50(E=1)}(x) \end{aligned}$$

This assumption implies tighter bounds above the median in the income distribution:

$$\begin{aligned}
F\left(w \mid x, E=1\right) P(x) &\leq F\left(w \mid x\right) \\
&\leq F\left(w \mid x, E=1\right) P(x) + \left[1 - P(x)\right] \qquad \text{if } w < w^{50(E=1)}(x) \\
& F\left(w \mid x, E=1\right) P(x) + 0.5\left[1 - P(x)\right] \\
&\leq F\left(w \mid x\right) \\
&\leq F\left(w \mid x, E=1\right) P(x) + \left[1 - P(x)\right] \qquad \text{if } w \geq w^{50(E=1)}(x)
\end{aligned}$$

2.4 The Roy Model

The economy consists of two occupations: fishers and hunters. Each individual can choose between these occupations.

$Y_{1i}-$ Earnings if individual i chooses to be a hunter.

$Y_{2i}-$ Earnings if individual i chooses to be a fisher.

The individuals are not equally productive in these occupations and can expect different earnings.

$Y_{1i} = \mu_1 + u_{1i}$

$Y_{2i} = \mu_2 + u_{2i}$

(Y_1, Y_2) joint normal distribution with covariance matrix $\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$.

$u_1 = Y_1 - \mu_1$

$u_2 = Y_2 - \mu_2$

$Z = \frac{\mu_1 - \mu_2}{\sigma}$ and $u = \frac{u_2 - u_1}{\sigma}$.

$\sigma^2 = V\left(u_1 - u_2\right)$

The individual chooses to be a hunter if

$$\begin{aligned} Y_1 &> Y_2 \\ \mu_1 - \mu_2 &> u_2 - u_1 \\ Z &> u \end{aligned}$$

The expected *observed* earnings for hunters will then be

$$E\left(Y_1 | u < Z\right) = \mu_1 - \sigma_{1u} \frac{\phi\left(Z\right)}{\Phi\left(Z\right)},$$

where $\sigma_{1u} = cov\left(u_1, u\right)$.

The expected *observed* earnings for fishers is

$$E\left(Y_2 | u > Z\right) = \mu_2 + \sigma_{2u} \frac{\phi\left(Z\right)}{1 - \Phi\left(Z\right)},$$

where $\sigma_{2u} = cov\left(u_2, u\right)$.

Case 1:
$\sigma_{1u} < 0$
$\sigma_{2u} > 0.$

Observed incomes for both fishers and hunters are above μ_1 and μ_2 respectively. Individuals choose occupation according to their comparative advantage.

$E(Y_1|u < Z) > \mu_1$
$E(Y_2|u > Z) > \mu_2$

Case 2:
$\sigma_{1u} < 0$
$\sigma_{2u} < 0.$

High u_2 draws corresponds to low $u_2 - u_1$ draws.

This means that hunters are better in both hunting and fishing. Those who choose fishing are below average in both hunting and fishing - but they are better in fishing.

$E(Y_1|u < Z) > \mu_1$
$E(Y_2|u > Z) < \mu_2$

Observed average earnings for hunters is higher than μ_1 and observed average earnings for fishers are below μ_2.

2.5 The Type II Tobit Model

$$\begin{aligned} y_i^* &= \mathbf{x}_{1i}'\boldsymbol{\beta}_1 + \varepsilon_{1i} \\ h_i^* &= \mathbf{x}_{2i}'\boldsymbol{\beta}_2 + \varepsilon_{2i} \\ y_i &= y_i^*,\ h_i = 1 \text{ if } h_i^* > 0 \\ y_i, \text{ not observed } h_i &= 0 \text{ if } h_i^* \leqslant 0 \end{aligned}$$

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & 1 \end{pmatrix} \right]$$

$$\begin{aligned} E\left(y_i | h_i = 1\right) &= \mathbf{x}_{1i}'\boldsymbol{\beta}_1 + E\left(\varepsilon_{1i} | h_i = 1\right) \\ &= \mathbf{x}_{1i}'\boldsymbol{\beta}_1 + E\left(\varepsilon_{1i} | \varepsilon_{2i} > -\mathbf{x}_{2i}'\boldsymbol{\beta}_2\right) = \\ &= \mathbf{x}_{1i}'\boldsymbol{\beta}_1 + \frac{\sigma_{12}}{\sigma_2^2} \frac{\phi\left(\mathbf{x}_{2i}'\boldsymbol{\beta}_2\right)}{\Phi\left(\mathbf{x}_{2i}'\boldsymbol{\beta}_2\right)} = \\ &= \mathbf{x}_{1i}'\boldsymbol{\beta}_1 + \sigma_{12}\lambda\left(\mathbf{x}_{2i}'\boldsymbol{\beta}_2\right) \end{aligned}$$

Heckman's two stage method:

1. Use a probit and estimate $\lambda_i = \frac{\phi\left(\mathbf{x}_{2i}'\boldsymbol{\beta}_2\right)}{\Phi\left(\mathbf{x}_{2i}'\boldsymbol{\beta}_2\right)}$.
2. Include λ_i in $y_i = \mathbf{x}_{1i}'\boldsymbol{\beta}_1 + \sigma_{12}\lambda_i + \eta_i$, where $\eta_i = \varepsilon_{1i} - E\left(\varepsilon_{1i} | \mathbf{x}_i, h_i = 1\right)$.

Classical paper: Heckman (1979) "Sample selection bias as a specification error" Econometrica, 47, 153-161.

2.6 Willis and Rosen (1979)

"Education and Self-Selection" classical study on returns to schooling. Uses the Roy model.

Model:

$$\begin{aligned} Y_{ij} &= y_j (X_i, \tau_i) \\ j &= 1, ...n \end{aligned}$$

Y_{ij} is life-time earnings for individual i in education level j. X_i is observable characteristics affecting earnings and τ_i is unobservables.

Value Function: V_{ij} value for individual i to obtain education level j.Present value of all income streams throughout the individual's life cycle.

$$\begin{aligned} V_{ij} &= g (Y_{ij}, Z_i, \omega_i) \\ i \text{ belongs to } j \text{ if } V_{ij} &= \max (V_{i1} ......V_{in}) ; \end{aligned}$$

Z_i is observable characteristics affecting choice of schooling and ω_i is unobservable characteristics affecting level of schooling. ω and τ assumed to have a bivariate normal distribution, i.e.,

$$(\tau, \omega) \sim F (\tau, \omega) .$$

Two education levels: high school (B) and college or more (A).
$V_{ai}-$ value function if the individual chooses college or more.
$V_{bi}-$ value function if the individual chooses high school.

$$V_{ai} = \int y_{ai} (t) \exp (-r_i t) \, dt = [\overline{y}_{ai} / (r_i - g_{ai})] \exp (-r_i S) ,$$

where $\overline{y}_{ai}$ is the earnings level, g_{ai} is the earnings growth rate, r_i the subjective discount rate and S the length of schooling.

$$V_{bi} = \int y_{bi} (t) \exp (-r_i t) \, dt = [\overline{y}_{bi} / (r_i - g_{bi})] .$$

Indicator function

$$I_i = \ln\left(V_{ai}/V_{bi}\right)$$

$$\begin{aligned} \Pr(\text{choose } A) &= \Pr(V_a > V_b) = \Pr(I > 0) \\ \Pr(\text{choose } B) &= \Pr(V_a \leqslant V_b) = \Pr(I \leqslant 0) \end{aligned}$$

The model contains five equations: one for the earnings level and one for earnings growth rate at each education level, i.e.,

$$\begin{aligned} \ln \overline{y}_{ai} &= X_i\beta_a + u_{1i}, \\ g_{ai} &= X_i\gamma_a + u_{2i} \\ \ln \overline{y}_{bi} &= X_i\beta_b + u_{3i}, \\ g_{bi} &= X_i\gamma_b + u_{4i} \\ r_i &= Z_i\delta + u_{5i}. \end{aligned}$$

Z_i is observable characteristics affecting choice of schooling through the effect on the discount rate. For example, children from poor backgrounds have often higher costs for obtaining higher education. Where u_{ji} is assumed to follow a multivariate normal distribution $N\left(\mathbf{0}, \mathbf{\Sigma}\right)$.

Reduced form - obtained through a linearization of the difference between the value functions:

$$\begin{aligned} I &= \alpha_0 + X\left[\alpha_1\left(\beta_a - \beta_b\right) + \alpha_2\gamma_a + \alpha_3\gamma_b\right] + \alpha_4 Z\delta + \\ \alpha_1\left(u_1 - u_3\right) + \alpha_2 u_2 + \alpha_3 u_3 + \alpha_5 u_5 &= \\ & W\pi - \epsilon \\ \Pr(A \text{ observed}) &= \Pr\left(W\pi > \epsilon\right) = F\left(\frac{W\pi}{\sigma_\epsilon}\right) \end{aligned}$$

Earnings equations can be estimated using Heckman's strategy for correcting for sample selection bias, i.e.,

$$\begin{aligned} E\left(\ln \overline{y}_a \mid I > 0\right) &= X\beta_a + \frac{\sigma_{1\epsilon}}{\sigma_\epsilon}\lambda_a \\ E\left(g_a \mid I > 0\right) &= X\gamma_a + \frac{\sigma_{2\epsilon}}{\sigma_\epsilon}\lambda_a \end{aligned}$$

Willis and Rosen stress that identification can be made on functional form, i.e., X and Z could be identical. However, they specify X to include ability indicators and Z family background variables.

2.7 Björklund-Moffit (1987)

An individual joins the program if

$$U\left(Y_i+\alpha_i-\phi_i\right)>U\left(Y_i\right)$$

α_i- individual reward of participating in the program.
ϕ_i- individual cost of participating.

The model:
$Y_i=X_i\beta+\alpha_iT_i+\epsilon_i$
$T_i=1$ if $T_i^*>0; T_i=0$ otherwise
$T_i^*=\alpha_i-\phi_i$
$\alpha_i=Z_i\delta+u_i$
$\phi_i=W_i\eta+\nu_i$

$E\left(\epsilon_i\right)=\sigma_\epsilon^2$
$E\left(u_i\right)=\sigma_u^2$
$E\left(\nu_i\right)=\sigma_\nu^2$
$E\left(\epsilon_i\nu_i\right)=\sigma_{\epsilon\nu}^2$
$E\left(\epsilon_iu_i\right)=\sigma_{\epsilon u}^2$

$E\left(\nu_iu_i\right)=\sigma_{\nu u}^2$

Reduced form:

$Y_i=X_i\beta+Z_i\delta+\epsilon_i+u_i$, if $T_i=1$
$Y_i=X_i\beta+\epsilon_i$, if $T_i=0$
$T_i=1$ if $T_i^*>0; T_i=0$ otherwise
$T_i^*=Z_i\delta-W_i\eta+u_i-\nu_i$
Test for heterogeneity of rewards: $\sigma_u^2=\sigma_{\epsilon u}^2=\sigma_{u\nu}^2=0$ and $\delta=0$.

Test for heterogeneity of costs: $\sigma_\nu^2=\sigma_{\epsilon\nu}^2=\sigma_{u\nu}^2=0$ and $\eta=0$.

Requirement for identification: that at least one vector of Z_i and W_i is different

They are able to derive several interesting outcome measures:

The effect of treatment on the treated:
$E\left(\alpha_i|T_i=1,Z_i\delta,W_i\eta\right)=Z_i\delta+E\left(u_i|u_i-\nu_i>-Z_i\delta+W_i\eta\right)=$
$=Z_i\delta+\left(\sigma_{u,u-\nu}/\sigma_{u-\nu}\right)\lambda_i$

The marginal effect on rewards of changing the cost of the program:

$$\partial E\left(\alpha_i|T_i=1,Z_i\delta,W_i\eta\right)/\partial\left(W_i\eta\right)=\left[\sigma_{u,u-\nu}/\sigma^2_{u-\nu}\right]\lambda_i\left(\lambda_i-s_i\right)$$

Welfare effect:

$E\left(T_i^*|T_i=1,Z_i\delta,W_i\eta\right)=Z_i\delta-W_i\eta+E\left(u_i-\nu_i|u_i-\nu_i>-Z_i\delta+W_i\eta\right)=$
$=Z_i\delta-W_i\eta+\sigma_{u-\nu}\lambda_i$
where $\lambda_i=f\left(s_i\right)/\left[1-F\left(s_i\right)\right],\ s_i=\left(-Z_i\delta+W_i\eta\right)/\sigma_{u,u-\nu}$

2.8 Garen (1984)

1. Binary choice model

$$\begin{aligned} y_i &= a_0 + b_0 x_{1i} + \epsilon_{0i} \quad \text{if} \quad z = 0 \\ y_i &= a_1 + b_1 x_{1i} + \epsilon_{1i} \quad \text{if} \quad z = 1 \end{aligned}$$

$$\epsilon_{ji} \sim N\left(0, \sigma_{jj'}\right)$$

z^* is a latent variable:

$$z^* = y\left(z = 1\right) - y(z = 0) - c = \pi_0 + \pi_1 x_1 + \pi_2 x_2 + \xi,$$

i.e., a three equation model where the disturbances are correlated.

2. Continuous choice model

System of equations:

$$\begin{aligned} y_i &= a_0 + b_0 x_{1i} + \epsilon_{0i} \quad \text{if} \quad z = 0 \\ y_i &= a_1 + b_1 x_{1i} + \epsilon_{1i} \quad \text{if} \quad z = 1 \\ &\ldots \\ &\ldots \\ &\ldots \\ y_i &= a_n + b_n x_{1i} + \epsilon_{ni} \quad \text{if} \quad z = n \end{aligned}$$

If n is large, this can be approximated by a linear function:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 z + \beta_3 x_1^2 + \beta_4 z^2 + \beta_5 z \cdot x_1 + \epsilon + \phi \cdot z \qquad (*)$$

where $\epsilon + \phi \cdot z$ is a linear approximation of the ϵ_j's.

$$\epsilon_i \sim N\left(0, \sigma_\epsilon^2\right), \qquad \phi_i \sim N\left(0, \sigma_\phi^2\right), \qquad cov\left(\epsilon_{ij}, \phi_i\right) = \sigma_{\epsilon\phi}.$$

Leads to a two-equation model:

$$z = \pi_0 + \pi_1 x_1 + \pi_2 x_2 + \eta.$$

The reduced for of (*) can be estimated as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 z + \beta_3 x_1^2 + \beta_4 z^2 + \beta_5 z \cdot x_1 + \beta_6 \widehat{\eta} + \beta_7 \widehat{\eta} \cdot z + \epsilon.$$

2.9 Garen (1988)

Compensating wage differentials again:

$$\ln y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 q_i + \beta_3 r_i + \epsilon_i + \phi_{1i} q_i + \phi_{2i} r_i$$

q_i- risk of fatal injury.
r_i- risk of non-fatal injury.

q_i and r_i can be correlated to different components of the error term:
ϵ_i- remember Duncan and Holmlund!
q_i and ϕ_{1i} - can be hypotized that those who are willing to take more risks have higher rewards of taking risks.

Offsetting in come and substitution effects: high ϕ_1 creates higher income, which, if safety is a normal good, creates an income effect for demand of higher safty.

First-order approximations to q and r :

$$\begin{aligned} q &= \pi_0 + \pi_1 x_1 + \pi_2 x_2 + \pi_3 z + \eta \\ r &= \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 z + \mu \end{aligned}$$

ϵ_i, ϕ_{1i} and ϕ_{2i} depend on η and μ therefore OLS will be biased.

The wage equation can be written as

$$\begin{aligned} E(y|q,r,x_1) &= \beta_0 + \beta_1 x_1 + \beta_2 q + \beta_3 r + E(\epsilon_i + \phi_1 q + \phi_2 r | q, r, x_1) = \\ &= \mathbf{R}\boldsymbol{\beta} + E(\boldsymbol{\epsilon} + \phi_1 q + \phi_2 r | q = \mathbf{X}\boldsymbol{\pi} + \boldsymbol{\eta}, r = \mathbf{X}\boldsymbol{\Delta} + \boldsymbol{\mu}, x_1, q, r) = \\ &= \mathbf{R}\boldsymbol{\beta} + E(\boldsymbol{\epsilon} + \phi_1 q + \phi_2 r | \boldsymbol{\eta} = q - \mathbf{X}\boldsymbol{\pi}, \boldsymbol{\mu} = r - \mathbf{X}\boldsymbol{\Delta}, x_1, q, r) \end{aligned}$$

$$\begin{aligned} & E(\boldsymbol{\epsilon} + \phi_1 q + \phi_2 r | \boldsymbol{\eta} = q - \mathbf{X}\boldsymbol{\pi}, \boldsymbol{\mu} = r - \mathbf{X}\boldsymbol{\Delta}, x_1, q, r) \\ = \; & E(\boldsymbol{\epsilon} | \eta, \boldsymbol{\mu}, \mathbf{X}) + E(\phi_1 | \eta, \boldsymbol{\mu}, \mathbf{X}) q + E(\phi_2 | \eta, \boldsymbol{\mu}, \mathbf{X}) r \end{aligned}$$

$$y = \mathbf{R}\boldsymbol{\beta} + \boldsymbol{\gamma}_1 \widehat{\eta} + \boldsymbol{\gamma}_2 \widehat{\mu} + \boldsymbol{\gamma}_3 \widehat{\eta} \cdot q + \boldsymbol{\gamma}_4 \widehat{\eta} \cdot q + \boldsymbol{\gamma}_4 \widehat{\mu} \cdot q + \boldsymbol{\gamma}_5 \widehat{\eta} \cdot r + \boldsymbol{\gamma}_6 \widehat{\mu} \cdot q + \theta$$