# MPhil Econometrics – Limited Dependent Variables and Selection

Francis J. DiTraglia

University of Oxford

Complied on 2020-02-08 at 19:25:38

# Housekeeping

| | |
|---|---|
| **Lecturer:** | Francis J. DiTraglia |
| **Email:** | francis.ditraglia@ox.ac.uk |
| **Course Materials:** | `http://ditraglia.com/teaching` |
| **Office:** | 2132 Manor Road Building |
| **Meetings:** | after lecture on Wednesdays & Fridays or by appointment |

## References

- **Wooldridge (2010) – *Econometric Analysis of Cross Section & Panel Data***
- Cameron & Trivedi (2005) – *Microeconometrics: Methods and Applications*
- Train (2009) – *Discrete Choice Methods with Simulation*

# Lecture #5 – Sample Selection

Examples of Sample Selection

The Heckman Selection Model

# What is sample selection?

## Question

Thus far we have always assumed that $(y_1, \mathbf{x}_1), \ldots, (y_N, \mathbf{x}_N)$ are a random sample from the population of interest. What if they aren't?

## Example 1: MPhil Admissions

- Suppose we want to improve admissions decisions at Oxford.
- $y \equiv$ overall marks in 1st year of Oxford Economics MPhil
- $\mathbf{x} \equiv \{\text{undergrad grades, letters of reference}, \ldots\}$
- What we observe: $\mathbf{x}$ for all applicants; $y$ for applicants who were admitted.
- What we want: $\mathbb{E}(y|\mathbf{x})$ for all applicants.

# Example 2: A Model of Wage Offers

Gronau (1974; JPE)

## Question

How do wage offers offers $w_i^o$ vary with $\mathbf{x}_i$ for all people in the population.

## Problem

Only observe $w_i^o$ for people who *accept* their offer, i.e. those who are employed.

## Mathematically

$\mathbb{E}(w_i^o|\mathbf{x}_i) \neq \mathbb{E}(w_i^o|\mathbf{x}_i, \text{Employed})$

# The Heckman Selection Model (Heckit) — Is $\beta_1$ identified?

### Outcome Equation

$$y_1 = \mathbf{x}_1'\beta_1 + u_1$$

### Participation Equation

$$y_2 = \mathbb{1}\left\{\mathbf{x}'\delta_2 + v_2 > 0\right\}$$

### Assumptions

(a) Observe $y_2, \mathbf{x}' = (\mathbf{x}_1', \mathbf{x}_2')$; only observe $y_1$ if $y_2 = 1$.

(b) $(u_1, v_2)$ are mean zero and jointly independent of $\mathbf{x}$.

(c) $v_2 \sim \text{Normal}(0, 1)$

(d) $\mathbb{E}(u_1|v_2) = \gamma_1 v_2$ where $\gamma_1$ is an unknown constant.

### Notes

- $\mathbb{E}(u_1) = \mathbb{E}(v_2) = 0$ is not restrictive: just include intercepts in both equations.
- Assumption (d) would be *implied* by assuming that $(u_1, v_2)$ are jointly normal.
- These assumptions are strong. They can be weakened a bit, but not too much.

# Step 1: Show that $u_1$ and $\mathbf{x}$ are conditionally independent given $v_2$.

## Assumption (b)

$(u_1, v_2)$ are jointly independent of $\mathbf{x}$.

## Equivalently

$$f_{1,2|\mathbf{x}}(u_1, v_2|\mathbf{x}) = f_{1,2}(u_1, v_2), \quad \text{and} \quad f_{1|\mathbf{x}}(u_1|\mathbf{x}) = f_1(u_1), \quad \text{and} \quad f_{2|\mathbf{x}}(v_2|\mathbf{x}) = f_2(v_2)$$

## Therefore

$$f_{1|2,\mathbf{x}}(u_1|v_2, \mathbf{x}) = \frac{f_{1,2|\mathbf{x}}(u_1, v_2|\mathbf{x})}{f_{2|\mathbf{x}}(v_2|\mathbf{x})} = \frac{f_{1,2}(u_1, v_2)}{f_2(v_2)} = f_{1|2}(u_1|v_2)$$

## In Words

Conditioning on $(v_2, \mathbf{x})$ gives the same information about $u_1$ as conditioning on $v_2$ only.

# Step 2: Calculate $\mathbb{E}(y_1|\mathbf{x}, v_2)$; show that if $v_2$ were observed we'd be done.

$$
\begin{aligned}
\mathbb{E}(y_1|\mathbf{x}, v_2) &= \mathbb{E}(\mathbf{x}_1'\boldsymbol{\beta}_1 + u_1|\mathbf{x}, v_2) && \text{(Substitute Outcome Eq.)} \\
&= \mathbf{x}_1'\boldsymbol{\beta}_1 + \mathbb{E}(u_1|\mathbf{x}, v_2) && \text{($\mathbf{x}_1$ is a subset of $\mathbf{x}$)} \\
&= \mathbf{x}_1'\boldsymbol{\beta}_1 + \mathbb{E}(u_1|v_2) && \text{(apply result of Step 1)} \\
&= \mathbf{x}_1'\boldsymbol{\beta}_1 + \gamma_1 v_2 && \text{(apply Assumption (d))}
\end{aligned}
$$

This slide has been cut, but I'm leaving a blank slide in its place so that the slide numbering remains unchanged from lecture.

# Step 3: Relate $v_2$ (unobserved) to $\mathbf{x}$ and $y_2$ (both observed).

$$
\begin{aligned}
\mathbb{E}(y_1|\mathbf{x}, y_2) &= \mathbb{E}_{v_2|(\mathbf{x},y_2)}\left[\mathbb{E}\left(y_1|\mathbf{x}, y_2, v_2\right)\right] && \text{(Law of Iterated Expectations)} \\
&= \mathbb{E}_{v_2|(\mathbf{x},y_2)}\left[\mathbb{E}\left(y_1|\mathbf{x}, v_2\right)\right] && \text{(Participation Eq: } y_2 = g(\mathbf{x}, v_2)\text{)} \\
&= \mathbb{E}\left[\mathbf{x}_1'\boldsymbol{\beta}_1 + \gamma_1 v_2|\mathbf{x}, y_2\right] && \text{(apply result of Step 2)} \\
&= \mathbf{x}_1'\boldsymbol{\beta}_1 + \gamma_1 \mathbb{E}\left(v_2|\mathbf{x}, y_2\right) && (\mathbf{x}_1 \text{ is a subset of } \mathbf{x})
\end{aligned}
$$

Therefore

$$\mathbb{E}\left(y_1|\mathbf{x}, y_2 = 1\right) = \mathbf{x}_1'\boldsymbol{\beta}_1 + \gamma_1 \mathbb{E}(v_2|\mathbf{x}, y_2 = 1).$$

# What is the significance of Step 3?

- Define $h(\mathbf{x}) \equiv \mathbb{E}(v_2|\mathbf{x}, y_2 = 1)$. Then: $\mathbb{E}(y_1|\mathbf{x}, y_2 = 1) = \mathbf{x}_1'\boldsymbol{\beta}_1 + \gamma_1 h(\mathbf{x})$
- Note that $h(\mathbf{x})$ is a random variable: a function of $\mathbf{x}$.
- Step 3 shows that a linear regression of $y_1$ on $\mathbf{x}_1$ and $h(\mathbf{x})$ for the *selected* sample, those with $y_2 = 1$, identifies $\boldsymbol{\beta}_1$ and $\gamma_1$!
- All that remains is to figure out what function $h$ is...

# Note: Selection Bias Enters Through $\gamma_1$

## Assumption (d)

$\mathbb{E}(u_1|v_2) = \gamma_1 v_2$ allows *dependence* between errors in participation and outcome eqs.

## Step 3

$\mathbb{E}(y_1|\mathbf{x}, y_2 = 1) = \mathbf{x}_1'\boldsymbol{\beta}_1 + \gamma_1 \mathbb{E}(v_2|\mathbf{x}, y_2 = 1)$

## Therefore

If $\gamma_1 = 0$ there is no selection bias: in this case $\mathbb{E}(y_1|\mathbf{x}, y_2 = 1) = \mathbf{x}_1'\boldsymbol{\beta}$ so regressing $y_1$ on $\mathbf{x}_1$ for the subset of individuals with $y_2 = 1$ identifies $\boldsymbol{\beta}_1$.

## Step 4: Determine the distribution of $v_2$ given $(\mathbf{x}, y_2 = 1)$.

$$\mathbb{P}(v_2 \leq t|\mathbf{x}, y_2 = 1) = \mathbb{P}(v_2 \leq t|\mathbf{x}, v_2 > -\mathbf{x}'\boldsymbol{\delta}_2) \qquad \text{(participation eq.)}$$

$$= \frac{\mathbb{P}\left(\{v_2 \leq t\} \cap \{v_2 > -\mathbf{x}'\boldsymbol{\delta}_2\}|\mathbf{x}\right)}{\mathbb{P}(v_2 > -\mathbf{x}'\boldsymbol{\delta}_2|\mathbf{x})} \qquad \text{(defn. of cond. prob.)}$$

$$= \frac{\mathbb{P}\left\{v_2 \in (-\mathbf{x}'\boldsymbol{\delta}_2, t]\right\}}{\mathbb{P}(v_2 > -\mathbf{x}'\boldsymbol{\delta}_2)} \qquad (v_2 \text{ and } \mathbf{x} \text{ are indep.})$$

$$= \frac{\mathbb{P}\left\{z \in (c, t]\right\}}{\mathbb{P}(z > c)} \qquad (v_2 \text{ is standard normal})$$

where $z \sim \text{Normal}(0, 1)$ and we define the shorthand $c \equiv -\mathbf{x}'\delta_2$.

# Step 5: Calculate the Expectation of a Truncated Normal

Recall: $z \sim \text{Normal}(0,1)$ and $c \equiv -\mathbf{x}'\boldsymbol{\delta}_2$

CDF

$$\mathbb{P}(z \leq t | z > c) = \frac{\mathbb{P}\left\{z \in (c, t]\right\}}{\mathbb{P}(z > c)} = \mathbb{1}\left\{c \leq t\right\}\left[\frac{\Phi(t) - \Phi(c)}{1 - \Phi(c)}\right]$$

Density

$$f(z|z > c) = \frac{d}{dt}\mathbb{P}(z \leq t | z > c) = \begin{cases} 0, & z < c \\ \varphi(z)/\left[1 - \Phi(c)\right], & z \geq c \end{cases}$$

# Step 5: Calculate the Expectation of a Truncated Normal

Recall: $z \sim \text{Normal}(0,1)$ and $c \equiv -\mathbf{x}'\boldsymbol{\delta}_2$

$$\mathbb{E}(z|z>c) = \int_{-\infty}^{\infty} z f(z|z>c)\,dz = \frac{1}{1-\Phi(c)}\int_c^{\infty} z\varphi(z)\,dz$$

$$= \left[\frac{1}{1-\Phi(c)}\right]\left(\frac{1}{\sqrt{2\pi}}\right)\int_c^{\infty} z\exp\left\{-z^2/2\right\}\,dz$$

$$= \left[\frac{1}{1-\Phi(c)}\right]\left(\frac{1}{\sqrt{2\pi}}\right)\left[-\exp\left\{-z^2/2\right\}\right]_c^{\infty}$$

$$= \left[\frac{1}{1-\Phi(c)}\right]\left(\frac{\exp\left\{-c^2/2\right\}}{\sqrt{2\pi}}\right) = \frac{\varphi(c)}{1-\Phi(c)}$$

# Step 6: Put everything together.

## Recall: Step 3

$$y_1 = \mathbf{x}_1'\boldsymbol{\beta}_1 + \gamma_1 h(\mathbf{x}) + \eta, \quad \mathbb{E}\left[\eta | \mathbf{x}_1, h(\mathbf{x})\right] = 0, \quad h(\mathbf{x}) \equiv \mathbb{E}(v_2|\mathbf{x}, y_2 = 1)$$

## Using Steps 4–5

$$h(\mathbf{x}) = \frac{\varphi(-\mathbf{x}\boldsymbol{\delta}_2)}{1 - \Phi(-\mathbf{x}\boldsymbol{\delta}_2)} = \frac{\varphi(\mathbf{x}'\boldsymbol{\delta}_2)}{\Phi(\mathbf{x}'\boldsymbol{\delta}_2)} \quad \text{since } \varphi(-c) = \varphi(c) \text{ and } 1 - \Phi(c) = \Phi(-c).$$

## Inverse Mills Ratio

$\varphi(c)/\Phi(c)$ is the inverse Mills Ratio, traditionally denoted by $\lambda \implies h(\mathbf{x}) = \lambda(\mathbf{x}'\boldsymbol{\delta}_2)$.

## Careful!

In an earlier lecture $\lambda$ denoted the standard logistic density. Here it's something else!

# The Heckman Two-step Estimator aka "Heckit"

## Observables

Observe $(y_{2i}, \mathbf{x}_i)$ for a random sample of size $N$; only observe $y_{1i}$ for those with $y_{2i} = 1$.

## First Step – Estimate $\delta_2$ from Full Sample

- Run Probit on the Participation Eq. $\mathbb{P}(y_{2i} = 1|\mathbf{x}_i) = \Phi(\mathbf{x}_i'\boldsymbol{\delta}_2)$ for the full sample.
- Define $\widehat{\lambda}_i \equiv \lambda(\mathbf{x}_i'\widehat{\boldsymbol{\delta}}_2)$ where $\widehat{\boldsymbol{\delta}}_2$ is the MLE for $\boldsymbol{\delta}_2$.

## Second Step – Estimate $(\beta_1, \gamma_1)$ from Selected Sample

Using the observations for which $y_{i1}$ is observed, regress $y_{i1}$ on $(\mathbf{x}_{1i}, \widehat{\lambda}_i)$ by OLS to obtain estimates $(\widehat{\boldsymbol{\beta}}_1, \widehat{\gamma}_1)$.

# The Heckman Two-step Estimator aka "Heckit"

## Theorem

Under the assumptions from above, the 2-step "Heckit" estimators satisfy

$$\begin{bmatrix} \widehat{\boldsymbol{\delta}}_2 \\ \widehat{\boldsymbol{\beta}}_1 \\ \widehat{\gamma}_1 \end{bmatrix} \rightarrow_p \begin{bmatrix} \boldsymbol{\delta}_2 \\ \boldsymbol{\beta}_1 \\ \gamma_1 \end{bmatrix} \quad \text{and} \quad \sqrt{N}\begin{bmatrix} \widehat{\boldsymbol{\delta}}_2 - \boldsymbol{\delta}_2 \\ \widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \widehat{\gamma}_1 - \gamma_1 \end{bmatrix} \rightarrow_d \text{Normal}(\mathbf{0}, \boldsymbol{\Omega}) \quad \text{as } N \to \infty.$$

## Standard Errors

The asymptotic variance matrix $\boldsymbol{\Omega}$ is complicated: the usual OLS standard errors from step two are incorrect as they do not account for the estimation of $\boldsymbol{\delta}_2$ in step one.

# The Big Picture: How does Heckit solve the selection problem?

- If we regress $y_{1i}$ on $\mathbf{x}_{1i}$ for the selected sample, there is an omitted variable.
- Under the Heckit assumptions, the omitted variable is precisely $\lambda(\mathbf{x}_i' \boldsymbol{\delta}_2)$.
- Hence: a regression of $y_{1i}$ on $\mathbf{x}_{1i}$ and $\lambda(\mathbf{x}_i' \boldsymbol{\delta}_2)$ is correctly specified.

# Why is the second step regression identified?

- If $\mathbf{x}_i$ contains some variables that are *not* in $\mathbf{x}_{1i}$, we have an exclusion restriction: i.e. there are variables that affect participation but not outcomes.
- Even if there are no exclusion restrictions, $\lambda$ is nonlinear so $\lambda(\mathbf{x}_{1i}'\boldsymbol{\delta}_2)$ will not be perfectly co-linear with $\mathbf{x}_{1i}$.
- Without exclusion restrictions identification comes *solely* from nonlinearity in $\lambda$.
- Depending on the values where it is evaluated, $\lambda$ can be *close* to linear, leading to very imprecise estimates unless you have an exclusion restriction.