

# Selectivity Models and Censored Models

Li and Racine (2007, Chapters 10 and 11)

---

Yasuyuki Matsumura

January 5, 2025

Graduate School of Economics, Kyoto University

# Sample Selection Issues

---

# Sample Selection

- Sample selection issues frequently arise in empirical studies.
- We concern that the treatment effect for those “selected as treated” will differ from that for persons randomly selected from the general population.
- Pioneering parametric approaches to deal with sample selection can be found in Heckman (1976, 1979).

# **Semiparametric Type-2 Tobit Models**

---

## Type-2 Tobit Models

- The Type-2 Tobit model is the four equation system:

$$Y_{1i}^* = X_{1i}^T \beta_1 + u_{1i},$$

$$Y_{2i}^* = X_{2i}^T \beta_2 + u_{2i},$$

$$Y_{1i} = 1(Y_{1i}^* > 0),$$

$$Y_{2i} = Y_{2i}^* \times 1(Y_{1i} = 1).$$

- The variables  $(Y_{1i}^*, Y_{2i}^*)$  are latent (unobserved).
- The observed variables are  $(Y_{1i}, Y_{2i}, X_{1i}, X_{2i})$ .
- Effectively,  $Y_{2i}^*$  is only when  $Y_{1i} = 1$ , equivalently when  $Y_{1i}^* > 0$ .
- Typically, the second equation is of interest, e.g. the coefficient  $\beta_2$ .

## Type-2 Tobit Models: Estimation

- The Type-2 Tobit model is a classical selection model introduced by Heckman (1976).
- It is conventional to assume that the error terms  $(u_{1i}, u_{2i})$  are independent of  $X_i = (X_{1i}, X_{2i})$ .
- For details of Heckman's estimation procedure, see the attached pdf file.
- Heckman's estimation is one of the parametric approaches, as he imposes the following parametric distributional assumptions on the joint distribution of the errors:

$$\begin{bmatrix} u_{1i} \\ u_{2i} \end{bmatrix} \sim \text{Normal} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \right)$$

## Semiparametric Type-2 Tobit Models

- Here we do not impose parametric distributional assumptions on the joint distribution of the errors.
- Assume that  $(u_{1i}, u_{2i})$  are independent of  $X_i = (X_{1i}, X_{2i})$ .
- Then, we obtain

$$\begin{aligned}\mathbb{E}(Y_{2i} \mid X_i, Y_{1i} = 1) &= X_{2i}^T \beta_2 + \mathbb{E}(u_{2i} \mid X_i, Y_{1i} = 1) \\ &\equiv X_{2i}^T \beta_2 + g(X_{1i}^T \beta_1),\end{aligned}$$

where

$$g(z) = \mathbb{E}(u_{2i} \mid u_{1i} > -z) = 1 - F_{u_2|u_1}(-z),$$

and  $F_{u_2|u_1}(\cdot)$  is the conditional CDF of  $u_{2i}$  given  $u_{1i}$ .

- The functional form of  $g(\cdot)$  is unknown.

- The simple regression of  $Y_{2i}$  on  $X_{2i}$  using the available data yields

$$Y_{2i} = X_{2i}^T \beta_2 + g(X_{1i}^T \beta_1) + \epsilon_{2i},$$
$$\mathbb{E}(\epsilon_{2i} \mid X_i, Y_{1i} = 1) = 0,$$

which is a partially linear single index model.

- Here we review the following estimation methods:
  - Powell (1987),
  - Ichimura and Lee (1991),
  - Gallant and Nychka (1987),
  - Heckman (1990)<sup>1</sup>.

---

<sup>1</sup>He proposes an estimation of intercept.



- Powell (1987) proposes a two-step estimation procedure.
- **Infeasible Estimation:** Define  $Z_i = X_{1i}^T \beta_1$ . If  $Z_i$  were observed, the regression is

$$Y_{2i} = X_{2i}^T \beta_2 + g(Z_i) + \epsilon_{2i},$$

which is a partially linear model.

- This can be estimated using Robinson's approach.
- Note that the intercept is absorbed by  $g(\cdot)$ , and that it must be excluded from  $X_{2i}$ <sup>2</sup>.

---

<sup>2</sup>Recall the identification conditions in semiparametric partially linear models.

## Powell (1987): Feasible Two-Step Estimation

- In practice,  $Z_i$  is not observed.
- We implement the following two-step approach.
- **1st Step:** Estimate  $\beta_1$  by a semiparametric binary choice estimator, or by Powell's (1984) CLAD estimator <sup>3</sup>.
- **2nd Step:** Let  $\hat{\beta}_1$  denote the estimator of  $\beta_1$ .
- Replace  $Z_i$  with  $\hat{Z}_i = X_{1i}^T \hat{\beta}_1$ :  $Y_{2i} = X_{2i}^T \beta_2 + g(\hat{Z}_i) + \epsilon_{2i}$ .
- Estimate  $\beta_2$  and  $g(\cdot)$  by Robinson's estimator.
- Note that  $\hat{Z}_i = X_{1i}^T \hat{\beta}_1$  is a generated regressor, and that the asymptotic distribution will differ from the result presented in Chapter 7.

---

<sup>3</sup>CLAD 推定量を使うのは、後ほど触れる Type-3 Tobit の 2 段階推定の方が適切なのでは？

## Ichimura and Lee (1991)

- Ichimura and Lee (1991) propose a joint estimator for  $\theta = (\beta_1^T, \beta_2^T)^T$  based on the nonlinear regression:

$$Y_{2i} = X_{2i} + g(X_{1i}\beta_1) + \epsilon_{2i}$$

for observations  $i$  such that  $Y_{2i}$  is observed.

- Their objective function is given by

$$Q_n(\theta) = \frac{1}{n} \sum_{i=1}^n 1(X_i \in \mathcal{X}) [Y_i - X_{2i}^T \beta_2 - \hat{g}(X_{1i}^T \beta_1)]^2,$$

where

$$\hat{g}(X_{1i}^T \beta_1) = \frac{\sum_{j \neq i} (Y_{2j} - X_{2j}^T \beta_2) K_h \left( \frac{(X_{1i} - X_{1j})^T \beta_1}{h} \right)}{\sum_{j \neq i} K_h \left( \frac{(X_{1i} - X_{1j})^T \beta_1}{h} \right)}$$

is a leave-one-out NW estimator of  $\mathbb{E}(Y_{2i} - X_{2i}^T \beta_2 \mid X_{1i}^T \beta_1)$ <sup>4</sup>.

<sup>4</sup>教科書は leave-one-out になっていない。

- Their estimator is an extension of a NLLS Heckit estimator, which is based on the equation

$$Y_{2i} = X_{2i}^T \beta_2 + \sigma_{12} \lambda(X_{1i}^T \beta_1) + \epsilon_{2i}.$$

- Such estimators ignore the first equation in the system.
- This is convenient as it simplifies the estimation.
- However, ignoring relevant information reduces efficiency.
- Ichimura and Lee derive the asymptotic normality:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \text{Normal}(0, A^{-1} \Sigma A^{-1}).$$

- They also derive consistent estimators  $\hat{A}$  and  $\hat{\Sigma}$ .

## Gallant and Nychka (1987): Semi-Nonparametric MLE

- Gallant and Nychka suggest approximating the joint density of the error terms  $f(u_{1i}, u_{2i})$  by a series expansion:

$$\tilde{f}(u_{1i}, u_{2i}) = \exp \left[ -\frac{u_{1i}^2}{2\sigma_1^2} - \frac{u_{2i}^2}{2\sigma_2^2} \right] \left[ \sum_{j=0}^K \sum_{k=0}^K \gamma_{jk} u_{1i}^j u_{2i}^k \right].$$

- $\tilde{f}(u_{1i}, u_{2i})$  is a baseline distribution of a joint normal expression.
- This is accompanied by a power series expansion allowing for a general form of the CDF.
- Using the above joint density formula, we can compute  $\tilde{f}(u_{1i}, u_{2i})$  and then construct a log-likelihood function.
- Maximizing the log-likelihood function, we obtain estimators of  $\beta_1$  and other parameters.
- The estimator has consistency under  $K \rightarrow \infty$ , and  $\frac{K}{n} \rightarrow 0$  as  $n \rightarrow \infty$ .

- Coppejans and Gallant (2002) show that one can use data-driven methods to select the power series expansion terms when estimating  $f(u_{1i}, u_{2i})$ .
- Newey (1999) proposes a two-step series-based estimation method.
  - First, estimate  $\beta_1$  efficiently.
  - Second, select  $\beta_2$  solving an efficient score equation.
- For details of nonparametric series methods, see Li and Racine (2007, Chapter 15).
- 教科書にタイポが多いのと、元論文にアクセスできないのとで、不正確な内容が含まれているかもしれないため、何かおかしい点があれば指摘していただきたいです。

## Heckman (1990): Intercept Estimation

- In the semiparametric Type-2 Tobit model, we cannot identify an intercept term, which cannot be separated from  $g(\cdot)$ .
- One might be interested in the intercept, for example, when determining “wage gaps” between unionized and non-unionized workers, or when decomposing wage differentials between different socioeconomic groups, etc  $\dots$ .

- Letting  $\mu$  denote the intercept, we write

$$Y_{2i}^* = \mu + \tilde{X}_{2i}^T \delta + u_{2i},$$

where  $X_{2i} = (1, \tilde{X}_{2i}^T)^T$ , and  $\beta_2 = (\mu, \delta^T)^T$ .

- Then, we obtain

$$\mathbb{E}(Y_{2i} \mid X_i, Y_{1i} = 1) = \mu + \tilde{X}_{2i}^T \delta + g(X_{1i}^T \beta_1).$$

- Recall the definition of  $g(\cdot)$ :

$$g(z) = \mathbb{E}(u_{2i} \mid u_{1i} > -z),$$

which leads to

$$\lim_{z \rightarrow \infty} g(z) = \mathbb{E}(u_{2i}) = 0, \text{ or}$$

$$\lim_{z \rightarrow \infty} \mathbb{E}(Y_{2i} - \tilde{X}_{2i}^T \delta \mid Y_{1i} = 1, X_{1i}^T \beta_1 > z) = \mu.$$



- Heckman (1990) suggests to use the observations such that  $\mathbb{E}(u_{2i} \mid Y_{1i} = 1) = g(X_{1i}^T \beta_1)$ , i.e., the observations such that  $g(\cdot)$  satisfies  $g(-\infty) = 0$ .
- Thus,  $\mu$  can be rewritten as

$$\mu = \mathbb{E}(Y_{2i} - \tilde{X}_{2i}^T \delta \mid Y_{1i} = 1, X_{1i}^T \beta_1 > \gamma_n),$$

where  $\gamma \rightarrow \infty$  is a bandwidth.

- This can be estimated by

$$\tilde{\mu} = \frac{\sum_{i=1}^n (Y_{2i} - \tilde{X}_{2i}^T \hat{\delta}) Y_{1i} 1(X_{1i}^T \hat{\beta}_1 > \gamma_n)}{\sum_{i=1}^n Y_{1i} 1(X_{1i}^T \hat{\beta}_1 > \gamma_n)}.$$

## Extension: Andrews and Schafgans (1998)

- Since the indicator function  $1(\cdot)$  is not differentiable, it is difficult to examine the asymptotic distribution of  $\tilde{\mu}$ .
- Andrews and Schafgans (1998) suggest to replace the indicator function with a smoothed non-decreasing CDF  $s(\cdot)$ , which satisfies

$$s(z) = 0 \text{ for } z \leq 0,$$

$$s(z) = 1 \text{ for } z \geq b \text{ for some } 0 < b < \infty, \text{ and}$$

$$s(\cdot) \text{ has third bounded derivatives.}$$

- They estimate  $\mu$  by

$$\hat{\mu} = \frac{\sum_{i=1}^n (Y_{2i} - \tilde{X}_{2i}\hat{\delta})Y_{1i}s(X_{1i}^T\hat{\beta}_1 > \gamma_n)}{\sum_{i=1}^n Y_{1i}s(X_{1i}^T\hat{\beta}_1 > \gamma_n)}.$$

- They find that the asymptotic distribution has a non-standard rate, depending on the distribution of  $X_{1i}^T\beta_1$ .

# **Semiparametric Type-3 Tobit Models**

---

## Type-3 Tobit Models

- The Type-3 Tobit model is the four equation system:

$$Y_{1i}^* = X_{1i}^T \beta_1 + u_{1i},$$

$$Y_{2i}^* = X_{2i}^T \beta_2 + u_{2i},$$

$$Y_{1i} = \max\{Y_{1i}^*, 0\},$$

$$Y_{2i} = Y_{2i}^* \times 1(Y_{1i} > 0).$$

- The difference from the Type-2 is that  $Y_{1i}$  is censored than binary.
- We observe  $Y_{2i}^*$  only when there is no censoring on  $Y_{1i}^*$ .
- Typically, the second equation is of interest, e.g. the coefficient  $\beta_2$ .

## Type-3 Tobit Models: Parametric Approaches

- Parametric approaches to estimate the Type-3 Tobit models impose parametric distributional assumptions on the joint distribution of the errors.
- Vella (1992, 1998)
- Wooldridge (1994)
- See the attached file for details.

## Semiparametric Type-3 Tobit Models

- Here we do not assume that the joint distribution of  $(u_{1i}, u_{2i})$  is known.
- Instead, we have  $\mathbb{E}(u_{2i} \mid u_{1i}) = g(u_{1i})$ , where  $g(\cdot)$  is an unknown function.
- In this case, it is easy to see that  $\mathbb{E}(Y_{2i} \mid X_i, u_{1i}) = X_{2i}^T \beta_2 + g(u_{1i})$ .
- Thus, we obtain

$$Y_{2i} = X_{2i}^T \beta_2 + g(u_{1i}) + v_{2i},$$

$$\mathbb{E}(v_{2i} \mid u_{1i}, Y_{1i} > 0) = 0.$$

## Semiparametric Type-3 Tobit Models: Estimation

- If  $u_{1i}$  were known, this would be a partially linear model.
- In practice,  $u_{1i}$  is unknown.
- We can estimate  $u_{1i}$  by Tobit, CLAD, etc  $\dots$ .
- Here we review the following estimation methods:
  - Li and Wooldridge (2002),
  - Chen (1997),
  - Honore, Kyriazidou and Udry (1997),
  - Lee (1994),
  - the semiparametric Type-2 Tobit estimator based on Ichimura (1993); and Ichimura and Lee (1991).

- Li and Wooldridge suggest a multistep method to estimate  $\beta_2$ .
- **1st Step:** Estimate  $\beta_1$ , for example, by Powell's (1984) CLAD estimator<sup>5</sup>. We assume that for the 1st step there is a  $\sqrt{n}$ -consistent, and asymptotically normally distributed, estimator for  $\beta_1$ .
- **2nd Step:** Replacing  $u_{1i}$  with  $\hat{u}_{1i}$ , we implement Robinson's approach to estimate  $\beta_2$ .
- They establish the  $\sqrt{n}$ -normality of their estimator  $\hat{\beta}_2$ :

$$\sqrt{n}(\hat{\beta}_2 - \beta_2) \xrightarrow{d} \text{Normal}(0, \Sigma).$$

- $\text{Avar}(\hat{\beta}_2)$  can be consistently estimated.

<sup>5</sup>Powell's CLAD (censored least absolute deviation) estimator will be reviewed below in Chapter 11.



- Li and Wooldridge's (2002) estimator does not achieve the semiparametric efficiency bound. Efficient estimation can usually be achieved by a one-step procedure, where  $\beta_1$  and  $\beta_2$  are estimated simultaneously as in Ai (1997).
- Powell's CLAD estimator is a parametric approach, i.e., the generated regressor is estimated from a parametric model. Ahn and Powell (1993) suggest to estimate  $\beta_1$  in the 1st stage using a nonparametric regression model.

- Assume that  $(U_{1i}, u_{2i})$  is independent of  $(X_{1i}, X_{2i})$ .
- Under the above assumption, it holds that

$$\begin{aligned}\mathbb{E}(Y_{2i} \mid X_{1i}, X_{2i}, u_{1i} > 0, X_{1i}^T \beta_1 > 0, Y_{1i} > 0) \\ &= \mathbb{E}(Y_{2i} \mid u_{1i} > 0, X_i) \\ &= X_{2i}^T \beta_2 + \alpha_0,\end{aligned}$$

where  $\alpha_0$  is a constant. Note that  $\alpha_0$  is not the intercept of the original model.

- **1st Step:** Estimate  $\beta_1$  consistently by Honore and Powell (1984), or by Powell's CLAD. Let  $\hat{\beta}_1$  denote the consistent estimator of  $\beta_1$ .
- **2nd Step:** Run the following least squares:

$$\min_{\beta_2, \alpha} \frac{1}{n} \sum_{i=1}^n 1\{Y_{1i} - X_{1i}^T \hat{\beta}_1 > 0, X_{1i}^T \hat{\beta}_1 > 0\} (Y_{2i} - X_{2i}^T \beta_2 - \alpha)^2.$$

- A problem arising with Chen's (1997) estimator is that it may trim out too many observations, which leads to inefficient estimation.
- Chen (1997) suggests an alternative estimator that trims far fewer data points in finite-sample applications.

## Honore, Kyriazidou and Udry (1997)

- Honore, Kyriazidou and Udry (1997) suggest to relax the normality assumption as it can be seen in Heckman (1979).
- Instead, they assume that the distribution of the error terms  $(u_{1i}, u_{2i})$  given regressors  $X_i$  is symmetric, with arbitrary heteroscedasticity permitted <sup>6</sup>.
- In this case, conditional on the sample selection, the conditional distribution is no longer symmetric.
- Their basic idea is that  $u_{2i}$  is symmetrically distributed around 0 if one estimates  $\beta_2$  using observations for which  $-X_{1i}^T \beta_1 < u_{1i} < X_{1i}^T \beta_1$  (i.e.,  $0 < Y_{1i} < 2X_{1i}^T \beta_1$ ).

---

<sup>6</sup>Note that the symmetry condition is neither weaker nor stronger than the independence condition.

- Honore, Kyriazidou and Udry (1997) suggest the following estimation method.
- **1st Step:** Estimate  $\beta_1$  consistently, for example, by Powell's CLAD. Let  $\hat{\beta}_1$  denote the consistent estimator of  $\beta_1$ .
- **2nd Step:** Run the following least absolute deviations:

$$\min_{\beta_2} \frac{1}{n} \sum_{i=1}^n 1\{0 < Y_{1i} < 2X_{1i}^T \hat{\beta}_1\} | Y_{2i} - X_{2i}^T \beta_2 | .$$

- They establish the  $\sqrt{n}$ -normality of their estimator.

Under the assumption of independence between the errors and the regressors, Lee (1994, equation 2.12) shows that

$$\begin{aligned} Y_{2i} - E(Y_2|u_1 > -X'_{1i}\beta_1, X'_1\beta > X'_{1i}\beta_1) \\ = [X'_{2i} - E(X'_2|X'_1\beta_1 > X'_{1i}\beta_1)]\beta_2 + u_{2i}, \end{aligned} \quad (10.25)$$

where  $u_{2i}$  satisfies  $E(u_{2i}|u_1 > -X'_{1i}\beta_1, X'_1\beta > X'_{1i}\beta_1) = 0$ . Lee suggests first replacing the conditional expectations in (10.25) by kernel estimators (also  $\beta_1$  needs to be replaced by a first stage estimator) and then applying a least squares procedure to estimate  $\beta_2$  (which we denote by  $\hat{\beta}_2$ , Lee). Lee establishes the asymptotic normality of  $\hat{\beta}_2$ , Lee.

## Comparing the 4 Estimators

	LW	Chen	HKU	Lee
Kernel Methods	Required	-	-	Required
Smoothing Parameter Choice	Insensitive	-	-	Insensitive

- In general, nonparametric kernel methods are sensitive to the choice of smoothing parameters.
- Lee (1994); Min, Sheu and Wang (2003) suggest by Monte-Carlo simulations that the estimators of LW(2002) and Lee (1994) are fairly insensitive to the choice of smoothing parameters.
- The reason is that the semiparametric estimators depend on the average of nonparametric estimators, which are less sensitive to different values of smoothing parameters than a pointwise nonparametric kernel estimator.

	LW	Chen	HKU	Lee
Dependence Assumption	Yes	Yes	No	Yes
Symmetry Assumption	No	No	Yes	No

- “Dependence Assumption” means that one need to assume that  $(u_{1i}, u_{2i})$  is independent of  $(X_{1i}, X_{2i})$ .
- “Symmetry Assumption” means that one need to assume that the distribution of the error terms  $(u_{1i}, u_{2i})$  given regressors  $X_i$  is symmetric.
- The symmetry condition is neither weaker nor stronger than the independence condition.



## Tests for Selection Bias

---

# Tests for Selection Bias

# **Nonparametric Sample Selection Model**

---



## References

---

- Textbooks and Lecture Notes:
  - Li, Q., and J. S. Racine (2007). *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.
  - 末石直也 (2015) 『計量経済学：ミクロデータ分析へのいざない』 日本評論社.
  - 末石直也 (2024) 『データ駆動型回帰分析：計量経済学と機械学習の融合』 日本評論社.
  - 西山慶彦, 新谷元嗣, 川口大司, 奥井亮 (2019) 『計量経済学』 有斐閣.
  - ECON 718 NonParametric Econometrics (Bruce Hansen, Spring 2009, University of Wisconsin-Madison).
  - セミノンパラメトリック計量分析 (末石直也, 2014 年度後期, 京都大学).