Econometrics II TA Session #5

Hiroki Kato

1 Empirical Application of Truncated Regression: Labor Participation of Married Women

1.1 Background and Data

To develop women's social advancement, we should create environment to keep a good balance between work and childcare after marriage. In this application, using the dataset of married women, we explore how much childcare prevents married women to participate in labor market.

Our dataset originally comes from Stata sample data. [1] This dataset contains the following variables:

- `whrs`: Hours of work. This outcome variable is truncated from below at zero.
- `kl6`: the number of preschool children
- `k618`: The number of school-aged children
- `wa`: age
- `we`: The number of years of education

```
dt <- read.csv(file = "./data/labor.csv", header = TRUE,  sep = ",")
summary(dt)
```

```
##       whrs           kl6              k618             wa
##  Min.   :  12   Min.   :0.0000   Min.   :0.000   Min.   :30.00
##  1st Qu.: 645   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:35.00
##  Median :1406   Median :0.0000   Median :1.000   Median :43.50
##  Mean   :1333   Mean   :0.1733   Mean   :1.313   Mean   :42.79
##  3rd Qu.:1903   3rd Qu.:0.0000   3rd Qu.:2.000   3rd Qu.:48.75
##  Max.   :4950   Max.   :2.0000   Max.   :8.000   Max.   :60.00
##        we
##  Min.   : 6.00
##  1st Qu.:12.00
##  Median :12.00
```

[1] http://www.stata-press.com/data/r13/laborsub.dt. Because this is dta file, we need to import it, using the `read.dta` function in the library `foreign`. I intentionally remove married women who could not participate in the labor market.

```
##  Mean   :12.64
##  3rd Qu.:13.75
##  Max.   :17.00
```

1.2 Model

Since we cannot observe those who could not partiapte in the labor market (`whrs = 0`), we use the truncated regression model. Thus, the selection rule is as follows:

$$\begin{cases} y_i = x_i\beta + u_i & \text{if } s_i = 1 \\ s_i = 1 & \text{if } a_1 < y_i < a_2 \end{cases}.$$

where $u_i \sim N(0, \sigma^2)$. In this case, $a_1 = 0$ and $a_2 = +\infty$.

Since we are interested in estimating β, we must condition on $s_i = 1$. The probability density function of y_i conditional on $(x_i, s_i = 1)$ is

$$p_\theta(y_i|x_i, s_i = 1) = \frac{f(y_i|x_i)}{\int_0^{+\infty} f(y_i|x_i)dy_i}.$$

where $\theta = (\beta, \sigma^2)'$. Because the distribution of y_i depends on the distribution of u_i, using $u_i = y_i - x_i\beta$, we obtain

$$p_\theta(u_i|x_i, -x_i\beta < u_i) = \frac{1}{\sigma}\frac{\phi(\frac{y_i - x_i\beta}{\sigma})}{1 - \Phi(\frac{-x_i\beta}{\sigma})}.$$

Thus, the log-likelihood function is

$$M_n(\theta) = \sum_{i=1}^{n} \log\left(\frac{1}{\sigma}\frac{\phi(\frac{y_i - x_i\beta}{\sigma})}{1 - \Phi(\frac{-x_i\beta}{\sigma})}\right).$$

We provide two ways to estimate truncated regression, using `R`. First way is to define the log-likelihood function directly and minimize its function by `nlm` function. Recall that `nlm` function provides the Newton method to minimize the function. We need to give intial values in argument of this function. Coefficients of explanatory variables, `b[3:6]`, are zero, and intercept, `b[2]`, and σ, `b[1]`, are given by mean and standard deviation of `whrs`, respectively.

```
whrs <- dt$whrs
kl6 <- dt$kl6; k618 <- dt$k618
wa <- dt$wa; we <- dt$we

LnLik <- function(b) {
  sigma <- b[1]
  xb <- b[2] + b[3]*kl6 + b[4]*k618 + b[5]*wa + b[6]*we
  condp <- dnorm((whrs - xb)/sigma)/(1 - pnorm(-xb/sigma))
```

```
  LL_i <- log(condp/sigma)
  LL <- -sum(LL_i)
  return(LL)
}

init <- c(sd(whrs), mean(whrs), 0, 0, 0, 0)
est.LnLik <- nlm(LnLik, init, hessian = TRUE)
```

Second way is to use the function `truncreg` in the library `truncreg`. This function must specify the trucated point in arguments `point` and `direction`. If `direction = "left"`, the outcome variable is truncated from below at `point`, that is, `point < y`. On the other hand, if `direction = "right"`, the outcome variable is truncated from above at `point`, that is, `y < point`.

```
library(truncreg)
model <- whrs ~ kl6 + k618 + wa + we
est.trunc <- truncreg(model, data = dt, point = 0, direction = "left")
se.trunc <- sqrt(diag(vcov(est.trunc)))
```

1.3 Interpretations

Table 1 shows results of truncated regression estimated by two methods. As a comparison, we also show the OLS result in column (3). All specifications show that the number of preschool and school-aged children reduces the hours of work. The size of coefficient of the number of preschool and school-aged children become stronger when we apply the truncated regression. Although the relationship between labor participation and women's characteristics is statistically insignificant, size of coefficients largely differs among three specifications.

```
ols <- lm(model, data = dt)
coef.LnLik <- est.LnLik$estimate
se.LnLik <- sqrt(diag(solve(est.LnLik$hessian)))
names(coef.LnLik) <- c("sigma", names(coef(ols)))
names(se.LnLik) <- c("sigma", names(coef(ols)))

library(stargazer)
stargazer(
  ols, ols, ols,
  column.labels = c("Truncated (truncreg)", "Truncated (nlm)", "OLS"),
  coef = list(coef(est.trunc), coef.LnLik[2:6]),
  se = list(se.trunc, se.LnLik[2:6]),
  report = "vcs", keep.stat = c("n"),
  covariate.labels = c(
    "\\#.Preschool Children",
    "\\#.School-aged Children",
    "Age", "Education Years"
```

Table 1: Truncated Regression: Labor Market Participation of Married Women

	Dependent variable:		
	whrs		
	Truncated (truncreg)	Truncated (nlm)	OLS
	(1)	(2)	(3)
#.Preschool Children	−456.785	−803.032	−421.482
	(266.367)	(252.803)	(167.973)
#.School-aged Children	−153.347	−172.875	−104.457
	(81.780)	(100.590)	(54.186)
Age	−5.379	−8.821	−4.785
	(13.492)	(14.646)	(9.691)
Education Years	−0.092	16.529	9.353
	(43.702)	(46.430)	(31.238)
Constant	1,624.584	1,586.228	1,629.817
	(857.730)	(932.878)	(615.130)
Estimated Sigma	941.464	983.736	
Log-Likelihood	-1201.698	-1200.916	
Observations	150	150	150

```
  ),
  add.lines = list(
    c("Estimated Sigma",
      round(coef(est.trunc)[6], 3), round(coef.LnLik[1], 3)),
    c("Log-Likelihood",
      round(est.trunc$logLik, 3), round(-est.LnLik$minimum, 3))
  ),
  omit.table.layout = "n", table.placement = "t",
  title = "Truncated Regression: Labor Market Participation of Married Women",
  label = "lfp",
  type = "latex", header = FALSE
)
```