

# Econometrics II TA Session #04 \*

Kenta KUDO †

November 12th, 2019

## Contents

1	Preliminary	2
2	Tobit Model	2
3	Count Data Model	3
3.1	Poisson Count Data Model . . . . .	3
3.2	Zero Inflated Poisson Count Data Model . . . . .	4
4	Empirical Examples	5
4.1	Empirical Example–Logit/Probit Model . . . . .	5
4.2	Empirical Example–Zero Inflated Poisson Count Data Model . . . . .	8
AppendixA	Poisson Distribution	10
AppendixB	Non-linear Optimization Procedure	10

---

\* All comments welcome!

† E-mail: vge011kk@student.econ.osaka-u.ac.jp

# 1 Preliminary

Today, we review the truncated regression model (Tobit model), the count data model and examples of the estimation.

2.1 Tobit Model

3.1 Count Data Model

3.2 Zero Inflated Poisson Count Data Model

4.1 Empirical Example–Logit/Probit Model

4.2 Empirical Example–Zero Inflated Poisson Count Data Model

A Poisson Distribution

B Non-linear Optimization Procedure

## 2 Tobit Model

The following discussion is explained in Hayashi (2000). Suppose that  $y_i$  and  $X_i$  satisfy following linear regression model such that:

$$y_i^* = X_i' \beta + u_i, \quad u_i \sim N(0, \sigma^2), \quad i = 1, 2, \dots, n \quad (1)$$

where  $y_i$  is i.i.d. dependent variable and  $X_i' \in \mathbb{R}^{1 \times k}$  is explanatory variable. Consider the following censoring:

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > c, \\ c & \text{if } y_i^* \leq c. \end{cases} \quad (2)$$

That is, an equivalent way to write the model is

$$y_i = \max\{X_i' \beta + u_i, c\}. \quad (3)$$

This model is called **the Tobit model** to pay respect to Tobin (1958). Unlike in truncated model, there is no truncation here. The feature that distinguishes the censored regression model from usual regression model is that the dependent variable is censored. For the observations that satisfy  $y_i^* > c$ , its density is given by

$$\sigma^{-1} \phi \left( \frac{y_i - X_i' \beta}{\sigma} \right). \quad (4)$$

Because there is no truncation, this density differs from the density for the truncated model in this case. For the observations which satisfy  $y_i^* \leq c$ , the probability is given by

$$\text{Prob}(y_i^* \leq c | X_i) = \text{Prob} \left( \frac{y_i^* - X_i' \beta}{\sigma} \leq \frac{c - X_i' \beta}{\sigma} \middle| X_i \right) \quad (5)$$

$$= \Phi \left( \frac{c - X_i' \beta}{\sigma} \right), \quad (6)$$

since we have  $(y_i^* - X_i'\beta)/\sigma | X_i \sim N(0, 1)$ . Therefore, the density of  $y_i$ , defined over the interval  $[c, +\infty)$ , is Eq. (4) for the case  $y_i > c$  and a probability mass of size  $\Phi\left(\frac{c - X_i'\beta}{\sigma}\right)$  at  $y_i = c$ . This density can be rewritten as

$$\left[\sigma^{-1}\phi\left(\frac{y_i - X_i'\beta}{\sigma}\right)\right]^{I_i} \times \left[\Phi\left(\frac{c - X_i'\beta}{\sigma}\right)\right]^{1-I_i}.$$

Recall that  $I_i$  is an indicator function such that:

$$I_i = \begin{cases} 1 & \text{if } y_i^* > c, \\ 0 & \text{if } y_i^* \leq c. \end{cases}$$

Taking the logarithms, we obtain the log-conditional likelihood for the observation  $i = 1, 2, \dots, n$  as follows:

$$\log L(\beta, \sigma^2) = I_i \log \left[\sigma^{-1}\phi\left(\frac{y_i - X_i'\beta}{\sigma}\right)\right] + (1 - I_i) \log \left[\Phi\left(\frac{c - X_i'\beta}{\sigma}\right)\right]. \quad (7)$$

### 3 Count Data Model

Like in the case of traffic accidents, we sometimes need to know how many times the events occur. In this section, we assume that the observed events are counted as  $y_i = 0, 1, 2, \dots$

#### 3.1 Poisson Count Data Model

Consider the following Poisson distribution:

$$\mathbb{P}(X = x) = f(x) := \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{for } x = 0, 1, 2, \dots$$

The **Poisson distribution** is obtained as a limit of a **binomial distribution** (二項分布) by  $n \rightarrow \infty$  and  $p \rightarrow 0$  with  $np = \lambda (\in \mathbb{R}_{++})$  fixed, where  $p$  stands for the success probability of a trial. The expectation of this distribution is

$$\mathbb{E}[X] = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=0}^{\infty} \lambda \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} = \lambda \underbrace{\sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}}_1 = \lambda,$$

Recall that  $\sum_{i=0}^{\infty} e^{-\lambda}\lambda^x/x! = 1$  since this represents the total sum of probability. Moreover, the variance is

$$\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \lambda(\lambda + 1) - \lambda^2 = \lambda.$$

Note that the second moment of  $X$  is calculated as follows:

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda}\lambda^x}{x!} = \lambda \sum_{x=1}^{\infty} x \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} = \lambda \left( \sum_{x^*=0}^{\infty} (x^* + 1) \frac{e^{-\lambda}\lambda^{x^*}}{x^*!} \right) \\ &= \lambda \mathbb{E}[x^* + 1] = \lambda(\lambda + 1). \end{aligned}$$

### 3.1.1 Model

Let  $y_i \in \{0, 1, 2, \dots\}$  (discrete numbers) and  $y_i \sim Poi(\lambda)$ . Poisson count data model is represented as

$$E[y_i] = \lambda_i = \exp(X_i\beta)$$

where  $\lambda_i > 0$ . ( $X_i\beta$  can be negative. So you should avoid the specification  $\lambda_i = X_i\beta$ .) To estimate  $\beta$ , we use the MLE method. The log-likelihood function is defined as

$$L(\beta) = f(y_1, y_2, \dots, y_n) = \prod_{i=1}^n f(y_i) = \prod_{i=1}^n \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!},$$

where  $\lambda_i = \exp(X_i\beta)$ . By taking natural logarithm, we obtain the following equation:

$$\begin{aligned} \log L(\beta) &= - \sum_{i=1}^n \lambda_i + \sum_{i=1}^n y_i \log \lambda_i - \sum_{i=1}^n \log(y_i!) \\ &= - \sum_{i=1}^n \exp(X_i\beta) + \sum_{i=1}^n y_i X_i\beta - \sum_{i=1}^n \log(y_i!) \end{aligned}$$

Then, by differentiating with  $\beta$ , we obtain the F.O.C. as follows:

$$\frac{\partial \log L(\beta)}{\partial \beta} = - \sum_{i=1}^n X_i' \exp(X_i\beta) + \sum_{i=1}^n X_i' y_i = 0.$$

Here we use the **Newton Raphson method** or the **Method of Scoring** to obtain  $\beta$ . We can apply the procedure as follows:

$$\beta^{(j+1)} = \beta^{(j)} - \left( - \sum_{i=1}^n X_i' X_i \exp(X_i\beta^{(j)}) \right)^{-1} \left( - \sum_{i=1}^n X_i' \exp(X_i\beta^{(j)}) + \sum_{i=1}^n X_i' y_i \right).$$

in the case of the Newton Raphson method, and

$$\beta^{(j+1)} = \beta^{(j)} - \mathbb{E} \left[ \left( - \sum_{i=1}^n X_i' X_i \exp(X_i\beta^{(j)}) \right)^{-1} \right] \left( - \sum_{i=1}^n X_i' \exp(X_i\beta^{(j)}) + \sum_{i=1}^n X_i' y_i \right).$$

in the case of the method of Scoring.

## 3.2 Zero Inflated Poisson Count Data Model

In the case of too many zeros, we should modify the Poisson count data model since the estimation does not become appropriate so often in such cases. So we assume that the probability of  $y_i = j$  is decomposed of two regimes as follows:

- **Regime 1:**  $y_i = 0$  *w.p.1*

Regime 1 is chosen with probability " $F(Z_i\alpha)$ ".

- **Regime 2:**  $y_i \sim Poi(\lambda_i)$

Regime 2 is chosen with probability “ $1 - F(Z_i\alpha)$ ”.

Therefore, the probability of observing  $y_i = j$  is

$$\mathbb{P}(y_i = 0) = \mathbb{P}(y_i = 0 | R_1)\mathbb{P}(R_1) + \mathbb{P}(y_i = 0 | R_2)\mathbb{P}(R_2), \quad (8)$$

$$\mathbb{P}(y_i = j) = \mathbb{P}(y_i = j | R_1)\mathbb{P}(R_1) + \mathbb{P}(y_i = j | R_2)\mathbb{P}(R_2). \quad (9)$$

In the above relationships, we assume

- $\mathbb{P}(y_i = 0|R_1) = 1$  and  $\mathbb{P}(y_i = j|R_1) = 0$  for  $j = 1, 2, \dots$ ;
- $\lambda_i = \exp(X_i\beta)$ ;
- $\mathbb{P}(R_1) = F(Z_i\alpha)$ ;
- $\mathbb{P}(R_2) = 1 - F(Z_i\alpha)$ ;
- $\mathbb{P}(y_i = j|R_2) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}$ .

Eq. (8) and Eq. (9) can be rewritten as

$$\mathbb{P}(y_i = j) = \mathbb{P}(R_1)I_i + \mathbb{P}(y_i = j | R_2)\mathbb{P}(R_2), \quad \text{for } j = 1, 2, \dots,$$

where  $I_i$  is an indicator variable which is equal to  $I_i = 1$  in the case of  $y_i = 0$  and  $I_i = 0$  in other cases. So probability function can be rewritten as

$$\mathbb{P}(y_i = j) = F(Z_i\alpha)I_i + \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}(1 - F(Z_i\alpha)), \quad j = 0, 1, 2, \dots$$

where the log-likelihood function is given by

$$\log L(\alpha, \beta) = \sum_{i=1}^n \log \mathbb{P}(y_i = j) = \sum_{i=1}^n \log \left( F(Z_i\alpha)I_i + \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}(1 - F(Z_i\alpha)) \right).$$

Then we obtained the estimator of  $\alpha$  and  $\beta$  by maximizing the log-likelihood function with respect to  $\alpha$  and  $\beta$  by means of the **Newton-Raphson Method** or the **Method of Scoring**.

## 4 Empirical Examples

### 4.1 Empirical Example-Logit/Probit Model

At first, we estimate the Logit model and the Probit model. The R code and data set are provided in the UCLA Institute for Digital Research and Education. <sup>\*1</sup> A researcher is interested in variables which indicates achievement degree, such as GRE (Graduate Record Exam scores), GPA (grade point average) and prestige of the undergraduate institution, effect admission into graduate school. The response variable, admit/do not admit is a binary variable. The model is represented as

$$\text{admit}_i = \text{GRE}_i\beta_1 + \text{GPA}_i\beta_2 + \text{rank}_i\beta_3 + u_i,$$

---

<sup>\*1</sup> <https://stats.idre.ucla.edu/>

where a variable "rank" represents the prestige of the university and takes on the values 1 through 4. The R code is given as follows.

```
#This example is provided in the UCLA Institute for Digital Research and
  Education
#https://stats.idre.ucla.edu/r/dae/logit-regression/

library(aod)
library(ggplot2)

mydata <- read.csv("https://stats.idre.ucla.edu/stat/data/binary.csv")
## view the first few rows of the data
head(mydata)

mydata$rank <- factor(mydata$rank)

#logit#####
mylogit <- glm(admit ~ gre + gpa + rank, data = mydata, family = "binomial
  ")
summary(mylogit)

#probit#####
myprobit <- glm(admit ~ gre + gpa + rank, family = binomial(link = "probit
  "),data = mydata)
summary(myprobit)
```

The result is explained as follows.

```
> library(aod)
Warning message:パッケージ “
  aod はバージョン 3.4.4 の R の下で造られました
> library(ggplot2)
> mydata <- read.csv("https://stats.idre.ucla.edu/stat/data/binary.csv")
> ## view the first few rows of the data
> head(mydata)
  admit gre  gpa rank
1     0 380 3.61   3
2     1 660 3.67   3
3     1 800 4.00   1
4     1 640 3.19   4
5     0 520 2.93   4
6     1 760 3.00   2
> mydata$rank <- factor(mydata$rank)
> mylogit <- glm(admit ~ gre + gpa + rank, data = mydata, family = "
  binomial")
>
> summary(mylogit)

Call:
glm(formula = admit ~ gre + gpa + rank, family = "binomial",
    data = mydata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6268  -0.8662  -0.6388   1.1490   2.0790
```

```

Coefficients:
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.989979   1.139951  -3.500 0.000465 ***
gre          0.002264   0.001094   2.070 0.038465 *
gpa          0.804038   0.331819   2.423 0.015388 *
rank2       -0.675443   0.316490  -2.134 0.032829 *
rank3       -1.340204   0.345306  -3.881 0.000104 ***
rank4       -1.551464   0.417832  -3.713 0.000205 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 499.98  on 399  degrees of freedom
Residual deviance: 458.52  on 394  degrees of freedom
AIC: 470.52

Number of Fisher Scoring iterations: 4

#####

Call:
glm(formula = admit ~ gre + gpa + rank, family = binomial(link = "probit"),
    data = mydata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6163  -0.8710  -0.6389   1.1560   2.1035

Coefficients:
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.386836   0.673946  -3.542 0.000398 ***
gre          0.001376   0.000650   2.116 0.034329 *
gpa          0.477730   0.197197   2.423 0.015410 *
rank2       -0.415399   0.194977  -2.131 0.033130 *
rank3       -0.812138   0.208358  -3.898 9.71e-05 ***
rank4       -0.935899   0.245272  -3.816 0.000136 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 499.98  on 399  degrees of freedom
Residual deviance: 458.41  on 394  degrees of freedom
AIC: 470.41

Number of Fisher Scoring iterations: 4

```

## 4.2 Empirical Example—Zero Inflated Poisson Count Data Model

Next, we estimate the zero-inflated Poisson count data model provided in the IDRE UCLA website.\*<sup>2</sup> This website explains a number of the examples introduced in the Econometrics II classes.

- We have data on 250 groups that went to a (state) park.
- Each group was questioned about how many fish they caught (count), how many children in the group (child), how many people were in the group (persons), and whether or not they brought a camper to the park (camper).
- Some visitors do not fish (Regime 1), but there is no data on whether a person fished or not.
- Some visitors who did fish (Regime 2) did not catch any fish so there are excess zeros in the data.

The R code is given as follows.

```
library(ggplot)
library(pscl)
library(stargazer)

zinb <- read.csv("https://stats.idre.ucla.edu/stat/data/fish.csv")
zinb <- within(zinb, {
  nofish <- factor(nofish)
  livebait <- factor(livebait)
  camper <- factor(camper)
})

summary(zinb)

summary(m1 <- zeroinfl(count ~ child + camper | persons, data = zinb))
stargazer(m1)
```

The result is explained as follows.

```
Call:
## zeroinfl(formula = count ~ child + camper | persons, data = zinb)
##
## Pearson residuals:
##      Min      1Q  Median      3Q      Max
## -1.2369 -0.7540 -0.6080 -0.1921 24.0847
##
## Count model coefficients (poisson with log link):
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.59789    0.08554  18.680 <2e-16 ***
## child       -1.04284    0.09999 -10.430 <2e-16 ***
## camper1      0.83402    0.09363   8.908 <2e-16 ***
##
```

---

\*<sup>2</sup> Institute for the Digital Research & Education "ZERO-INFLATED POISSON REGRESSION — R DATA ANALYSIS EXAMPLES" (<https://stats.idre.ucla.edu/r/dae/zip/>)



```
## Zero-inflation model coefficients (binomial with logit link):  
##           Estimate Std. Error z value Pr(>|z|)  
## (Intercept)  1.2974      0.3739   3.470 0.000520 ***  
## persons      -0.5643      0.1630  -3.463 0.000534 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Number of iterations in BFGS optimization: 12  
## Log-likelihood: -1032 on 5 Df
```

## Appendix

### AppendixA Poisson Distribution

Consider that an event occurs with probability  $p = \lambda/n$ , and we consider the probability that the event happens  $x$  times over  $n$  trials as follows:

$$P(X = x) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} = \frac{n!}{(n-x)!x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x},$$

where the distribution which  $X$  follows becomes a binomial distribution. Then, we can transform the above equation as follows:

$$P(X = x) = \frac{\overbrace{n \times (n-1) \times \dots \times (n-x+1)}^{x \text{ times multiplied}} \lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \quad (10)$$

$$= \frac{n}{n} \frac{(n-1)}{n} \frac{(n-2)}{n} \dots \frac{(n-x+1)}{n} \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x}. \quad (11)$$

By taking the limit of Eq. (11) with respect to  $n$ , with  $np = \lambda$  fixed (so that  $p \rightarrow 0$  accordingly), we obtain

$$\lim_{n \rightarrow \infty} P(X = x) = \lim_{n \rightarrow \infty} \underbrace{\frac{n}{n}}_1 \underbrace{\frac{(n-1)}{n}}_1 \dots \underbrace{\frac{(n-x+1)}{n}}_1 \frac{\lambda^x}{x!} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{e^{-\lambda}} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-x}}_1.$$

From the above equation, we have

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}.$$

### AppendixB Non-linear Optimization Procedure

Note that the Newton-Raphson method (one of the non-linear optimization procedure) is described as follows:

$$\beta^{(j+1)} = \beta^{(j)} - \left( \frac{\partial^2 \log L(\beta^{(j)})}{\partial \beta \partial \beta'} \right)^{-1} \frac{\partial \log L(\beta^{(j)})}{\partial \beta}$$

This equation comes from the first-order Taylor series expansion around  $\beta = \beta^*$ :

$$0 = \frac{\partial \log L(\beta)}{\partial \beta} \approx \frac{\partial \log L(\beta^*)}{\partial \beta} + \frac{\partial^2 \log L(\beta^*)}{\partial \beta \partial \beta'} (\beta - \beta^*)$$

Then we obtain

$$\frac{\partial^2 \log L(\beta^*)}{\partial \beta \partial \beta'} (\beta - \beta^*) = - \frac{\partial \log L(\beta^*)}{\partial \beta}$$
$$\beta - \beta^* = - \left( \frac{\partial^2 \log L(\beta^*)}{\partial \beta \partial \beta'} \right)^{-1} \frac{\partial \log L(\beta^*)}{\partial \beta}$$

This yields the above equation. If we take the expectation on the second derivative of the likelihood function, the method is known as the method of Scoring(スコア法).

## References

- [1] William, H., Greene (2012) "*Econometric analysis Seventh Edition*", Pearson.
- [2] Fumio, Hayashi (2000) "*ECONOMETRICS*", Princeton University Press.