2.3 Count Data Model (計数データモデル)

Poisson distribution:

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

for $x = 0, 1, 2, \cdots$.

In the case of Poisson random variable X, the expectation of X is:

$$\mathrm{E}(X) = \sum_{x=0}^{\infty} x\frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} x\frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} \lambda\frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} = \lambda\sum_{x'=0}^{\infty} \frac{e^{-\lambda}\lambda^{x'}}{x'!} = \lambda.$$

Remember that $\sum_x f(x) = 1$, i.e., $\sum_{x=0}^{\infty} e^{-\lambda}\lambda^x/x! = 1$.

Therefore, the probability function of the count data y_i is taken as the Poisson distribution with parameter λ_i.

In the case where the explained variable y_i takes 0, 1, 2, $\cdots$ (discrete numbers), assuming that the distribution of y_i is Poisson, the logarithm of λ_i is specified as a linear function, i.e.,

$$E(y_i) = \lambda_i = \exp(X_i\beta).$$

Note that λ_i should be positive.

Therefore, it is better to avoid the specification: $\lambda = X_i\beta$.

The joint distribution of $y_1, y_2, \cdots, y_n$ is:

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

where $\lambda_i = \exp(X_i\beta)$.

The log-likelihood function is:

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

The first-order condition is:

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

$\Longrightarrow$ Nonlinear optimization procedure

[Review] Nonlinear Optimization Procedures:

Note that the Newton-Raphson method (one of the nonlinear optimization procedures) is:

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

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

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

and β and β^* are replaced by $\beta^{(j+1)}$ and $\beta^{(j)}$, respectively.

An alternative nonlinear optimization procedure is known as the method of scoring, which is shown as:

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

where $\left(\dfrac{\partial^2 \log L(\beta^{(j)})}{\partial\beta\partial\beta'} \right)$ is replaced by $\mathrm{E}\left(\dfrac{\partial^2 \log L(\beta^{(j)})}{\partial\beta\partial\beta'} \right)$.

[End of Review]

In this case, we have the following iterative procedure:

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

The Newton-Raphson method is equivalent to the scoring method in this count model, because any random variable is not included in the expectation.

Zero-Inflated Poisson Count Data Model: In the case of too many zeros, we have to modify the estimation procedure.

Suppose that the probability of $y_i = j$ is decomposed of two regimes.

$\longrightarrow$ We have the case of $y_i = j$ and Regime 1, and that of $y_i = j$ and Regime 2.

Consider $P(y_i = 0)$ and $P(y_i = j)$ separately as follows:

$$P(y_i = 0) = P(y_i = 0|\text{Regime 1})P(\text{Regime 1}) + P(y_i = 0|\text{Regime 2})P(\text{Regime 2})$$

$$P(y_i = j) = P(y_i = j|\text{Regime 1})P(\text{Regime 1}) + P(y_i = j|\text{Regime 2})P(\text{Regime 2}),$$

for $j = 1, 2, \cdots$.

Assume:

- $P(y_i = 0|\text{Regime 1}) = 1$ and $P(y_i = j|\text{Regime 1}) = 0$ for $j = 1, 2, \cdots,$
- $P(\text{Regime 1}) = F_i$ and $P(\text{Regime 2}) = 1 - F_i,$
- $P(y_i = j|\text{Regime 2}) = \dfrac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}$ for $j = 0, 1, 2, \cdots,$

where $F_i = F(Z_i\alpha)$ and $\lambda_i = \exp(X_i\beta)$. $\Longrightarrow$ Z_i and X_i are exogenous variables.

Under the first assumption, we have the following equations:

$$P(y_i = 0) = P(\text{Regime 1}) + P(y_i = 0|\text{Regime 2})P(\text{Regime 2})$$
$$P(y_i = j) = P(y_i = j|\text{Regime 2})P(\text{Regime 2}),$$

for $j = 1, 2, \cdots$.

Combining the above two equations, we obtain the following:

$$P(y_i = j) = P(\text{Regime 1})I_i + P(y_i = j|\text{Regime 2})P(\text{Regime 2}),$$

for $j = 0, 1, 2, \cdots,$

where the indicator function I_i is given by $I_i = 1$ for $y_i = 0$ and $I_i = 0$ for $y_i \neq 0$.

F_i denotes a cumulative distribution function of $Z_i\alpha$. $\Longrightarrow$ We have to assume F_i.

Including the other two assumptions, we obtain the distribution of y_i as follows:

$$P(y_i = j) = F_i I_i + \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}(1 - F_i), \qquad j = 0, 1, 2, \cdots$$

where $F_i \equiv F(Z_i\alpha)$, $\lambda_i = \exp(X_i\beta)$, and the indicator function I_i is given by $I_i = 1$ for $y_i = 0$ and $I_i = 0$ for $y_i \neq 0$.

Therefore, the log-likelihood function is:

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

where $F_i \equiv F(Z_i\alpha)$ and $\lambda_i = \exp(X_i\beta)$.

$\log L(\alpha,\beta)$ is maximized with respect to α and β.

$\Longrightarrow$ The Newton-Raphson method or the method of scoring is utilized for maximization.

3 Panel Data

3.1 GLS — Review

Regression model I:

$$y = X\beta + u, \qquad\qquad u \sim N(0, \sigma^2 I_n),$$

where y, X, β, u, 0 and I_n are $n \times 1$, $n \times k$, $k \times 1$, $n \times 1$, $n \times 1$, and $n \times n$, respectively.

We solve the following minimization problem:

$$\min_{\beta} (y - X\beta)'(y - X\beta).$$

Let $\hat{\beta}$ be a solution of the above minimization problem.

OLS estimator of β is given by:

$$\hat{\beta} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'u.$$

$$\mathrm{E}(\hat{\beta}) = \beta, \qquad \mathrm{V}(\hat{\beta}) = \sigma^2(X'X)^{-1}.$$

Regression model II:

$$y = X\beta + u, \qquad u \sim N(0, \sigma^2\Omega),$$

where Ω is $n \times n$.

We solve the following minimization problem:

$$\min_{\beta} (y - X\beta)'\Omega^{-1}(y - X\beta).$$

Let b be a solution of the above minimization problem.

GLS estimator of β is given by:

$$b = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y = \beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}u.$$

$$\mathrm{E}(b) = \beta, \qquad \mathrm{V}(b) = \sigma^2(X'\Omega^{-1}X)^{-1}.$$

- We apply OLS to the following regression model:

$$y = X\beta + u, \qquad u \sim N(0, \sigma^2\Omega).$$

OLS estimator of β is given by:

$$\hat{\beta} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'u.$$

$$\mathrm{E}(\hat{\beta}) = \beta, \qquad \mathrm{V}(\hat{\beta}) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}.$$

$\hat{\beta}$ is an unbiased estimator.

The difference between two variances is:

$$
\begin{aligned}
& \mathrm{V}(\hat{\beta}) - \mathrm{V}(b) \\
&= \sigma^2 (X'X)^{-1} X'\Omega X (X'X)^{-1} - \sigma^2 (X'\Omega^{-1}X)^{-1} \\
&= \sigma^2 \Big((X'X)^{-1}X' - (X'\Omega^{-1}X)^{-1}X'\Omega^{-1} \Big) \Omega \Big((X'X)^{-1}X' - (X'\Omega^{-1}X)^{-1}X'\Omega^{-1} \Big)' \\
&= \sigma^2 A \Omega A'
\end{aligned}
$$

Ω is the variance-covariance matrix of u, which is a positive definite matrix.

Therefore, except for $\Omega = I_n$, $A\Omega A'$ is also a positive definite matrix.

This implies that $\mathrm{V}(\hat{\beta}_i) - \mathrm{V}(b_i) > 0$ for the ith element of β.

Accordingly, b is more efficient than $\hat{\beta}$.

3.2 Panel Model Basic

Model:

$$y_{it} = X_{it}\beta + v_i + u_{it}, \qquad i = 1, 2, \cdots, n, \quad t = 1, 2, \cdots, T$$

where i indicates individual and t denotes time.

There are n observations for each t.

u_{it} indicates the error term, assuming that $\mathrm{E}(u_{it}) = 0$, $\mathrm{V}(u_{it}) = \sigma_u^2$ and $\mathrm{Cov}(u_{it}, u_{js}) = 0$ for $i \neq j$ and $t \neq s$.

v_i denotes the individual effect, which is fixed or random.

3.2.1 Fixed Effect Model (固定効果モデル)

In the case where v_i is fixed, the case of $v_i = z_i\alpha$ is included.

$$y_{it} = X_{it}\beta + v_i + u_{it}, \qquad i = 1, 2, \cdots, n, \quad t = 1, 2, \cdots, T,$$

$$\overline{y}_i = \overline{X}_i\beta + v_i + \overline{u}_i, \qquad i = 1, 2, \cdots, n,$$

where $\overline{y}_i = \dfrac{1}{T}\displaystyle\sum_{t=1}^{T} y_{it}$, $\overline{X}_i = \dfrac{1}{T}\displaystyle\sum_{t=1}^{T} X_{it}$, and $\overline{u}_i = \dfrac{1}{T}\displaystyle\sum_{t=1}^{T} u_{it}$.

$$(y_{it} - \overline{y}_i) = (X_{it} - \overline{X}_i)\beta + (u_{it} - \overline{u}_i), \qquad i = 1, 2, \cdots, n, \quad t = 1, 2, \cdots, T,$$

Taking an example of y, the left-hand side of the above equation is rewritten as:

$$y_{it} - \overline{y}_i = y_{it} - \frac{1}{T} 1_T' y_i,$$

where $1_T = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$, which is a $T \times 1$ vector, and $y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}$.

$$\begin{pmatrix} y_{i1} - \overline{y}_i \\ y_{i2} - \overline{y}_i \\ \vdots \\ y_{iT} - \overline{y}_i \end{pmatrix} = I_T y_i - 1_T \overline{y}_i = I_T y_i - \frac{1}{T} 1_T 1_T' y_i = (I_T - \frac{1}{T} 1_T 1_T') y_i$$

Thus,

$$\begin{pmatrix} y_{i1} - \overline{y}_i \\ y_{i2} - \overline{y}_i \\ \vdots \\ y_{iT} - \overline{y}_i \end{pmatrix} = \begin{pmatrix} X_{i1} - \overline{X}_i \\ X_{i2} - \overline{X}_i \\ \vdots \\ X_{iT} - \overline{X}_i \end{pmatrix} \beta + \begin{pmatrix} u_{i1} - \overline{u}_i \\ u_{i2} - \overline{u}_i \\ \vdots \\ u_{iT} - \overline{u}_i \end{pmatrix}, \qquad i = 1, 2, \cdots, n,$$