Quantitative Analysis HW7

R11922045 陳昱行

1.

Let the expected value of Hessian matrix $H(\boldsymbol{\theta}_0) = \mathbb{E}[\nabla^2 \ln l(\boldsymbol{\theta}_0)]$

$$
\begin{gathered}
\mathbb{E}[\nabla^2 \ln l(\boldsymbol{\theta}_0)] \\
= \mathbb{E}\left[\nabla\left[\frac{[y_i - \Phi(x_i'\theta_0)]\phi(x_i'\theta_0)}{\Phi(x_i'\theta_0)[1-\Phi(x_i'\theta_0)]}x_i\right]\right] \\
= \mathbb{E}\left[\frac{\phi'(x_i'\theta_0)\Phi(x_i'\theta_0)y_i - \phi'(x_i'\theta_0)\Phi^2(x_i'\theta_0) - \phi'(x_i'\theta_0)\Phi^2(x_i'\theta_0)y_i + \phi'(x_i\theta_0)\Phi^3(x_i'\theta_0 \cdots}{(\Phi(x_i'\theta_0)[1-\Phi(x_i\theta_0)])^2}\right. \\
-\mathbb{E}\left[\frac{[y_i - 2y_i\Phi(x_i\theta_0) + \Phi^2(x_i'\theta_0)]\phi^2(x_i\theta_0)}{(\Phi(x_i'\theta_0)[1-\Phi(x_i\theta_0])^2}x_ix_i'\right] \\
= 0 - \mathbb{E}\left[\frac{[y_i - 2y_i\Phi(x_i\theta_0) + \Phi^2(x_i'\theta_0)]\phi^2(x_i\theta_0)}{(\Phi(x_i'\theta_0)[1-\Phi(x_i\theta_0])^2}x_ix_i'\right] \\
= -\mathbb{E}\left(\frac{\phi^2(x_i'\theta_0)}{\Phi(x_i'\theta_0)[1-\Phi(x_i'\theta_0)]}x_ix_i'\right)
\end{gathered}
$$

by law of iterated expectation and $\mathbb{E}(y_i|x_i) = \mathbb{P}(y_i = 1|x_i) = \Phi(x_i'\theta_0)$

which is the specification of the probit model.

For the information matirx $\boldsymbol{B}(\theta_0)$

$$
\boldsymbol{B}(\theta_0) = \mathbb{E}\left(\frac{1}{N}(\nabla L_n(\theta_0))(\nabla L_n(\theta_0))'\right)
$$

Since (y_i, x_i') are i.i.d data, the above expectation could be rewritten as

$$
\mathbb{E}\left(\frac{[y_i - \Phi(x_i'\theta_0)]^2\phi^2(x_i'\theta_0)}{(\Phi(x_i'\theta_0)[1-\Phi(x_i'\theta_0)])^2}x_ix_i'\right)
$$

By the property of binary variable:

$$
Var(y_i|x_i) = \Phi(x_i\theta_0)(1-\Phi(x_i'\theta_0))
$$

and definition of variance

$$
\begin{gathered}
Var(y_i|x_i) = \mathbb{E}(y_i^2|x_i) - \mathbb{E}^2(y_i|x_i) \\
= \mathbb{E}(y_i^2|x_i) - \Phi^2(x_i\theta_0)
\end{gathered}
$$

This implies

$$\mathbb{E}(y_i^2|x_i) = \Phi(x_i'\theta_0) = \mathbb{E}(y_i|x_i)$$

The above expectation could be futher reduce to

$$\mathbb{E}\left(\frac{\phi^2(x_i\theta_0)}{\Phi(x_i'\theta_0)[1-\Phi(x_i'\theta_0)]}x_i x_i'\right)$$

It follows that the information equality holds: $\boldsymbol{H}(\theta_0) + \boldsymbol{B}(\theta_0) = 0$

2.

For probit model, the NLS estimator is given by solving the following F.O.C:

$$\begin{aligned}&\frac{\partial}{\partial\theta}\sum_{i=1}^{n}[y_i - \Phi(x_i'\theta)]^2 \\ &= -2\sum_{i=1}^{n}[y_i - \Phi(x_i'\theta)]\phi(x_i\theta) = 0\end{aligned}$$

This is not the same as the ML method:

$$\sum_{i=1}^{n}\frac{y_i - \Phi(x_i'\theta)}{\Phi(x_i'\theta)[1-\Phi(x_i\theta)]}\phi(x_i'\theta)x_i = 0$$

For the logit model, the NLS estimator is given by solving the following F.O.C:

$$\begin{aligned}&\frac{\partial}{\partial\theta}\sum_{i=1}^{n}[y_i - G(x_i'\theta)]^2 \\ &= -2\sum_{i=1}^{n}[y_i - G(x_i'\theta)]G'(x_i'\theta)x_i = 0\end{aligned}$$

which is not the same as the ML method:

$$\sum_{i=1}^{n}[y_i - G(x_i'\theta)]x_i = 0$$

3.

(a)

The pdf of Bernoulli distribution is

$$f(y_i) = G(x_i'\theta)^{y_i}(1 - G(x_i\theta))^{1-y_i}$$

where $y_i = 0, 1.$

Therefore,

$$\begin{aligned}
L_n(\theta) &= \sum_{i=1}^{n}[y_i \ln G(x_i'\theta) + (1-y_i)\ln(1 - G(x_i'\theta))] \\
&= -\sum_{i=1}^{n} y_i \ln(1+e^{-x_i'\theta}) - \sum_{i=1}^{n}(1-y_i)\ln(1+e^{x_i'\theta}) \\
&= -\sum_{i=1}^{n}[y_i \ln(1+e^{-x_i'\theta}) - y_i \ln(1+e^{x_i'\theta}) + \ln(1+e^{x_i'\theta})] \\
&= -\sum_{i=1}^{n}[y_i \ln\left(\frac{1+e^{-x_i'\theta}}{1+e^{x_i'\theta}}\right) + \ln(1+e^{x_i'\theta})] \\
&= -\sum_{i=1}^{n}[-y_i(x_i'\theta) + \ln(1+e^{x_i'\theta})] \\
&= \sum_{i=1}^{n}[y_i(x_i'\theta) - \ln(1+e^{x_i'\theta})]
\end{aligned}$$

and $\nabla L_n(\theta)$ is defined by:

$$\begin{aligned}
\nabla L_n(\theta) &= \begin{bmatrix} \frac{\partial L_n(\theta)}{\partial \theta_0} \\ \frac{\partial L_n(\theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial L_n(\theta)}{\partial \theta_n} \end{bmatrix} \\
&= \sum_{i=1}^{n}[y_i - G(x_i\theta)]x_i
\end{aligned}$$

(b)

if $\boldsymbol{X}$ contains intercept and let $\tilde{\theta}$ be the solution such that $\nabla L_n(\tilde{\theta}) = 0$, then the above gradient could be written as:

$$
\begin{aligned}
&\sum_{i=1}^{n}[y_i - G(x_i\tilde{\theta})]\begin{bmatrix}1\\x_{i,1}\\x_{i,2}\\\vdots\end{bmatrix} = \mathbf{0}\\
\implies &\sum_{i=1}^{n}\begin{bmatrix}y_i - G(x_i'\tilde{\theta})\\(y_i - G(x_i'\tilde{\theta}))x_{i,1}\\(y_i - G(x_i'\tilde{\theta}))x_{i,2}\\\vdots\end{bmatrix} = \mathbf{0}\\
\implies &\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} G(x_i'\theta)
\end{aligned}
$$

from the first row of the matrix equality.