Suggested Solution to Problem Set 3

Prof. John Rust

Econ 551b, Spring 1999

QUESTION 1 Answers given in section 9 of Rust's lecture notes, Endogenous Regressors and Instrumental Variables.

QUESTION 2 Since

$$\varepsilon = y - X\beta \sim N(0, \Omega),$$

log likelihood function for β can be written as

$$l(\beta) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\Omega| - \frac{1}{2}\Big((y - X\beta)'\Omega^{-1}(y - X\beta)\Big).$$

MLE of β, $\hat{\beta}$, satisfies

$$\frac{\partial l(\hat{\beta})}{\partial \beta} = X'\Omega^{-1}(y - X\hat{\beta}) = 0.$$

Therefore

$$X'\Omega^{-1}y = X'\Omega^{-1}X\hat{\beta} \Rightarrow \hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y = \hat{\beta}_{GLS}.$$

QUESTION 3

$$\begin{aligned}
\mathcal{I}(\beta) &= E\left[\frac{\partial l(\hat{\beta})}{\partial \beta}\frac{\partial l(\hat{\beta})}{\partial \beta'}\right] \\
&= E\left[X'\Omega^{-1}(y - X\beta)(y - X\beta)'\Omega^{-1}X\right] \\
&= X'\Omega^{-1}E[(y - X\beta)(y - X\beta)']\Omega^{-1}X \\
&= X'\Omega^{-1}\Omega\Omega^{-1}X = X'\Omega^{-1}X
\end{aligned}$$

Therefore

$$\mathcal{I}(\beta)^{-1} = \Big(X'\Omega^{-1}X\Big)^{-1}.$$

QUESTION 4 Let N_k be the number of times which outcome k occured in all sample.

$$N_k = \sum_{i=1}^{N} I\{x_i = k\}, \qquad \sum_{k=1}^{K} N_k = N$$

With

$$f(x_i|\theta) = \prod_{k=1}^{K} \theta_k^{I\{x_i=k\}},$$

the joint distribution of $\{\tilde{X}_1, \ldots, \tilde{X}_N\}$ can be written as

$$f(x_1, \ldots, x_N|\theta) = \prod_{k=1}^{K} \theta_k^{N_k}.$$

1. Since log likelihood function is

$$\begin{aligned} l_N(\theta) &= \log f(x_1, \ldots, x_N|\theta) \\ &= \sum_{k=1}^{K} N_k \log \theta_k \\ &= \sum_{k=1}^{K-1} N_k \log \theta_k + \left(N - \sum_{k=1}^{K-1} N_k\right) \log\left(1 - \sum_{k=1}^{K-1} \theta_k\right), \end{aligned}$$

MLE $\widehat{\theta}_k$(for $k = 1, \ldots, K-1$) can be derived from

$$\frac{\partial}{\partial \theta_k} l_N(\theta) = \frac{N_k}{\theta_k} - \frac{(N - N_1 - \cdots - N_{K-1})}{(1 - \theta_1 - \cdots - \theta_{K-1})} = 0.$$

The solution to this system is

$$\widehat{\theta}_k = \frac{N_k}{N}.$$

2.Since

$$E\left[N_k\right] = \sum_{i=1}^{N} EI\{x_i = k\} = \sum_{i=1}^{N} \theta_k = N\theta_k$$

$$E\left[\widehat{\theta}_k\right] = E\left[\frac{N_k}{N}\right] = \frac{1}{N} E\left[N_k\right] = \frac{1}{N} N\theta_k = \theta_k.$$

Therefore $\widehat{\theta}_k$ is an unbiased estimator.
3.

$$\begin{aligned} E\left[N_k^2\right] &= E\left[\left(\sum_{i=1}^{N} I\{x_i = k\}\right)^2\right] \\ &= \sum_{i=1}^{N} EI\{x_i = k\}^2 + \sum_{i=1}^{N} \sum_{j \neq i}^{N} EI\{x_i = k\} I\{x_j = k\} \\ &= N\theta_k + N(N-1)\theta_k^2. \end{aligned}$$

$$E\left[\widehat{\theta}_k^2\right] = \frac{1}{N^2}E\left[N_k^2\right] = \frac{\theta_k}{N} + \frac{(N-1)\theta_k^2}{N}.$$

$$\begin{aligned}
\mathrm{var}(\widehat{\theta}_k) &= E\left[\widehat{\theta}_k^2\right] - \left\{E\left[\widehat{\theta}_k\right]\right\}^2 \\
&= \frac{\theta_k}{N} + \frac{(N-1)\theta_k^2}{N} - \theta_k^2 = \frac{1}{N}(1-\theta_k)\theta_k
\end{aligned}$$

For $k \neq l$

$$\begin{aligned}
E\left[N_k N_l\right] &= E\left[\left(\sum_{i=1}^{N} I\{x_i = k\}\right)\left(\sum_{i=1}^{N} I\{x_i = l\}\right)\right] \\
&= \sum_{i=1}^{N} EI\{x_i = k\}I\{x_i = l\} + \sum_{i=1}^{N}\sum_{j\neq i}^{N} EI\{x_i = k\}I\{x_j = l\} \\
&= 0 + \sum_{i=1}^{N}\sum_{j\neq i}^{N} EI\{x_i = k\}EI\{x_j = l\} \\
&= N(N-1)\theta_k\theta_l.
\end{aligned}$$

Note $E\left[I\{x_i = k\}I\{x_i = l\}\right] = 0$ since at least one of $I\{x_i = k\}$ and $I\{x_i = l\}$ must be zero.

$$E\left[\widehat{\theta}_k\widehat{\theta}_l\right] = \frac{1}{N^2}E\left[N_k N_l\right] = \frac{(N-1)\theta_k\theta_l}{N}$$

$$\begin{aligned}
\mathrm{cov}(\widehat{\theta}_k, \widehat{\theta}_l) &= E\left[\widehat{\theta}_k\widehat{\theta}_l\right] - E\left[\widehat{\theta}_k\right]E\left[\widehat{\theta}_l\right] \\
&= \frac{(N-1)\theta_k\theta_l}{N} - \theta_k\theta_l = -\frac{1}{N}\theta_k\theta_l
\end{aligned}$$

Therefore

$$\Sigma = \mathrm{cov}(\widehat{\theta}) = \frac{1}{N}\begin{bmatrix}
\theta_1(1-\theta_1) & -\theta_1\theta_2 & \cdots & -\theta_1\theta_{K-1} \\
-\theta_1\theta_2 & \theta_2(1-\theta_2) & \cdots & -\theta_2\theta_{K-1} \\
\cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots \\
-\theta_1\theta_{K-1} & \theta_2\theta_{K-1} & \cdots & \theta_{K-1}(1-\theta_{K-1})
\end{bmatrix}.$$

4.

For $k = l$

$$\frac{\partial^2}{\partial\theta_k^2} l_N(\theta) = -\frac{N_k}{\theta_k^2} - \frac{(N - N_1 - \cdots - N_{K-1})}{(1-\theta_1-\cdots-\theta_{K-1})^2}$$

$$\begin{aligned}
E\left[\frac{\partial^2}{\partial\theta_k^2} l_N(\theta)\right] &= -\frac{N\theta_k}{\theta_k^2} - \frac{(N - N\theta_1 - \cdots - N\theta_{K-1})}{(1-\theta_1-\cdots-\theta_{K-1})^2} \\
&= -\frac{N}{\theta_k} - \frac{N}{(1-\theta_1-\cdots-\theta_{K-1})}
\end{aligned}$$

For $k \neq l$

$$\frac{\partial^2}{\partial\theta_k \partial\theta_l} l_N(\theta) = -\frac{(N - N_1 - \cdots - N_{K-1})}{(1 - \theta_1 - \cdots - \theta_{K-1})^2}$$

$$\begin{aligned} E\left[\frac{\partial^2}{\partial\theta_k \partial\theta_l} l_N(\theta)\right] &= -\frac{(N - N\theta_1 - \cdots - N\theta_{K-1})}{(1 - \theta_1 - \cdots - \theta_{K-1})^2} \\ &= -\frac{N}{(1 - \theta_1 - \cdots - \theta_{K-1})} \end{aligned}$$

$$\begin{aligned} \mathcal{I}_N(\theta) &= -E\left[\frac{\partial^2}{\partial\theta \partial\theta'} l_N(\theta)\right] \\ &= N \begin{bmatrix} \frac{1}{\theta_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\theta_2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \frac{1}{\theta_{K-2}} & 0 \\ 0 & \cdots & \cdots & \frac{1}{\theta_{K-1}} \end{bmatrix} \\ &\quad + N \frac{1}{(1 - \theta_1 - \theta_2 - \cdots - \theta_{K-1})} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ 1 & \cdots & \cdots & 1 \\ 1 & \cdots & \cdots & 1 \\ 1 & 1 & \cdots & 1 \end{bmatrix}. \end{aligned}$$

5. MLE can be proved to be efficient if it has a variance equal to Cramer-Rao lower bound. You can easily show $\Sigma = \mathcal{I}_N^{-1}(\theta)$ by verifying $\Sigma \times \mathcal{I}_N(\theta) = I$.