Review Session: Bunche 3156, 10a - 11:30a, Tuesday

Final 2003, Q3

$U_{i,p} = X_{i,p}\beta_p + \varepsilon_{i,p}$ utility if stay

$U_{i,m} = X_{i,m}\beta_m + \varepsilon_{i,m}$ utility if migrate

$C_i = Z_i\gamma + u_i$ cost of migration

Migrate if $X_{i,m}\beta_m + \varepsilon_{i,m} - Z_i\gamma - u_i \geq X_{i,p}\beta_p + \varepsilon_{i,p}$ (1)

Define $y_i = \begin{cases} 1 & \text{if migrate} \\ 0 & \text{else} \end{cases}$

a)

$$\Pr[y_i = 1 \mid \underbrace{X_{i,m}, X_{i,p}, Z_i}_{\equiv W_i}] = \Pr[(1) \mid W_i]$$

$$= \Pr[\varepsilon_{i,m} - \varepsilon_{i,p} - u_i \geq X_{ip}\beta_p - X_{i,m}\beta_m + Z_i\gamma \mid W_i]$$

$$= \Pr[\varepsilon_{i,p} - \varepsilon_{i,m} + u_i \leq X_{i,m}\beta_m - X_{i,p}\beta_p - Z_i\gamma \mid W_i]$$

$$= F[X_{i,m}\beta_m - X_{i,p}\beta_p - Z_i\gamma \mid W_i]$$

where F is the cdf of $-\varepsilon_{i,p} + \varepsilon_{i,m} + u_i$ conditional on W_i

Since $\begin{bmatrix} \varepsilon_{i,p} \\ \varepsilon_{i,m} \\ u_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{pp} & \sigma_{pm} & \sigma_{pu} \\ \sigma_{pm} & \sigma_{mm} & \sigma_{mu} \\ \sigma_{pu} & \sigma_{mu} & \sigma_{uu} \end{bmatrix} \right)$

$$\Rightarrow [\,1 \;\; -1 \;\; 1\,] \begin{bmatrix} \varepsilon_{i,p} \\ \varepsilon_{i,m} \\ u_i \end{bmatrix} \sim N\left(0, [\sigma_{pp} - \sigma_{pm} + \sigma_{pu}, \; \sigma_{pm} - \sigma_{mm} + \sigma_{mu}, \; \sigma_{pu} - \sigma_{mu} + \sigma_{uu}] \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}\right)$$

$$\sim N(0, \sigma_{pp} - \sigma_{pm} + \sigma_{pu} - \sigma_{pm} + \sigma_{mm} - \sigma_{mu} + \sigma_{pu} - \sigma_{mu} + \sigma_{uu})$$

$$\sim N(0, \underbrace{\sigma_{pp} + \sigma_{mm} + \sigma_{uu} - 2\sigma_{pm} + 2\sigma_{pu} - 2\sigma_{mu}}_{\equiv \sigma^2})$$

$$\Rightarrow \Pr[y_i = 1 \mid W_i] = \Phi\left[\frac{X_{i,m}\beta_m - X_{i,p}\beta_p - Z_i\gamma}{\sigma} \mid W_i \right] \equiv \Phi_i$$

b) Construct the log-likelihood

$$L(\theta|W_i) = \prod_{i=1}^{n} \Phi_i^{Y_i} (1-\Phi_i)^{1-Y_i}$$

$$\Rightarrow \log L(\theta|W_i) = \sum_{i=1}^{n} \left[Y_i \ln \Phi_i + (1-Y_i) \ln (1-\Phi_i) \right]$$

c) $\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} \log L(\theta|W_i)$

<u>NLS:</u>

$$E[Y_i|W_i] = \Pr[Y_i = 1|W_i] = \Phi_i$$

$$\Rightarrow \hat{\theta}_{NLS} = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{n} (Y_i - \Phi_i)^2$$

Problem: $\eta_i \equiv Y_i - \Phi_i$ are heteroskedastic:

$$\eta_i = \begin{cases} 1-\Phi_i & \text{if } Y_i = 1 \text{ which is w/prob } \Phi_i \\ -\Phi_i & \text{if } Y_i = 0 \text{ which is w/prob } 1-\Phi_i \end{cases}$$

$$Var(\eta_i|W_i) = \Phi_i(1-\Phi_i)$$

$$\Rightarrow \hat{\theta}_{WNLS} = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{n} \frac{(Y_i - \Phi_i)^2}{\Phi_i(1-\Phi_i)}$$

Identification issue:

1] e.g. $X_{im} = [1 \quad race_i \quad unemp\ in\ NY_i]$

$X_{ip} = [1 \quad race_i \quad unemp\ in\ LA_i]$

$$\Rightarrow X_{im}\beta_m - X_{ip}\beta_p - Z_i\gamma = \beta_{0m} + \beta_{1m}\, black_i + \beta_{2m} NY_i - \beta_{0p} - \beta_{1p}\, black_i - \beta_{2p} LA_i - Z_i\gamma$$

Here, we can only estimate $\beta_{0m} - \beta_{0p}$ and $\beta_{1m} - \beta_{1p}$. Since the regressors NY_i and LA_i are different, we <u>can</u> identify β_{2m} and β_{2p}

2] Recall: $\Phi\left(\frac{X_{im}\beta_m - X_{ip}\beta_p - Z_i\gamma}{\sigma}\right)$

$\Rightarrow$ We can only estimate $\left(\frac{\beta_{0m}-\beta_{0p}}{\sigma}\right), \left(\frac{\beta_{1m}-\beta_{1p}}{\sigma}\right), \frac{\beta_{2m}}{\sigma}, \frac{\beta_{2p}}{\sigma}$

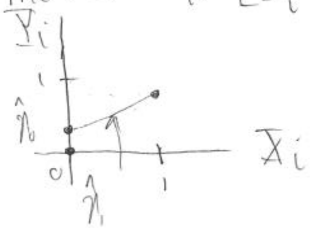
and $\frac{\gamma}{\sigma}$. (can only estimate up to scale)

d) $\left.\begin{matrix}\text{MLE}\\ \text{WNLS}\end{matrix}\right\} \xrightarrow{d} N(0, (I_1)^{-1})$

Final 2004, Q3

Logit model: $Pr(Y_i = 1 | X_i) = \Lambda(\beta_0 + \beta_1 X_i)$ where
$X_i \in \{0,1\}$

a) Show that this can be written as a linear probability model: $Pr[Y_i = 1 | X_i] = \gamma_0 + \gamma_1 X_i$



$$\left[\begin{aligned}\hat{\gamma}_0 &= \bar{Y}_i - \hat{\gamma}_1 \bar{X}_i \\ &= \frac{1}{n}\sum_{i=1}^{n} Y_i 1_{\{X_i=0\}} \\ \hat{\gamma}_1 &= \frac{1}{n}\sum_{i=1}^{n} Y_i 1_{\{X_i=1\}} - \frac{1}{n}\sum_{i=1}^{n} Y_i 1_{\{X_i=0\}}\end{aligned}\right. \qquad ? \quad \Big]$$

b) $\hat{\gamma}_0 = \bar{Y} - \hat{\gamma}_1 \bar{X}$

$$\hat{\gamma}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}$$

Note: $\sum X_i^2 = \sum X_i = n_1 \equiv \#$ times $X_i = 1$

$\bar{X} = \frac{1}{n}\sum X_i = \frac{n_1}{n}$

$\sum X_i Y_i = \sum_{j: X_j = 1} Y_j$

$$\hat{\gamma}_1 = \frac{\sum_{j:X_j=1} Y_j - n\frac{n_1}{n}\bar{Y}}{n_1 - n\frac{n_1^2}{n^2}} = \frac{\sum_{j:X_j=1} Y_j - n_1 \bar{Y}}{\underbrace{(n-n_1)n_1}_{n}}$$

define $n_0 \equiv n - n_1$

$$= \frac{n}{n_0 n_1} \sum_{j:X_j=1} Y_j - \frac{n}{n_0}\bar{Y}$$

but $n\bar{Y} = \sum_{i=1}^{n} Y_i$

$$= \frac{n}{n_0 n_1} \sum_{j:X_j=1} Y_j - \frac{1}{n_0}\left[\sum_{j:X_j=1} Y_j + \sum_{l:X_l=0} Y_l\right]$$

$$= \left(\frac{n}{n_0 n_1} - \frac{1}{n_0}\right) \sum_{j:X_j=1} Y_j - \underbrace{\frac{1}{n_0}\sum_{l:X_l=0} Y_l}_{\text{conditional average of } Y_l \text{ when } X_l=0}$$

$$= \frac{n-n_1}{n_0 n_1}$$

$$= \frac{n_0}{n_0 n_1} = \frac{1}{n_1}$$

$$= \frac{1}{n_1}\sum_{j:X_j=1} Y_j - \frac{1}{n_0}\sum_{l:X_l=0} Y_l$$

$$\Rightarrow \hat{\gamma}_0 = \frac{1}{n_0}\sum_{l:X_l=0} Y_l$$