MKT 7317 Problem Set 2

Jin Miao

jxm190071@utdallas.edu

University of Texas at Dallas

February 13, 2022

1 Nested Logit with Three Alternatives

Question (a)

Given $\rho = 1$, we have

$$F(\epsilon_1, \epsilon_2) = e^{-e^{-\epsilon_1} - e^{-\epsilon_2}}$$

$$\begin{cases} F_1(\epsilon_1) & = \lim_{\epsilon_2 \to \infty} F(\epsilon_1, \epsilon_2) = \lim_{\epsilon_2 \to \infty} e^{-e^{-\epsilon_1} - e^{-\epsilon_2}} = e^{-e^{-\epsilon_1}} \\ F_2(\epsilon_2) & = \lim_{\epsilon_1 \to \infty} F(\epsilon_1, \epsilon_2) = \lim_{\epsilon_1 \to \infty} e^{-e^{-\epsilon_1} - e^{-\epsilon_2}} = e^{-e^{-\epsilon_2}} \end{cases}$$

Given $F\epsilon_1, \epsilon_2 = F_1(\epsilon_1)F_2(\epsilon_2)$, we conclude that $\epsilon_1 \perp\!\!\!\perp \epsilon_2$.

We can also derive the probability distribution function as follows.

$$f(\epsilon_1, \epsilon_2) = \frac{\partial^2 F(\epsilon_1, \epsilon_2)}{\partial \epsilon_1 \partial \epsilon_2} = \epsilon_1 \epsilon_2 e^{-\epsilon_1} e^{\epsilon_2} e^{-e^{-\epsilon_1} - e^{-\epsilon_2}}$$

$$f_1(\epsilon_1) = \frac{dF_1(\epsilon_1)}{d\epsilon_1} = \epsilon_1 e^{-\epsilon_1} e^{-e^{-\epsilon_1}}$$

$$f_2(\epsilon_2) = \frac{dF_2(\epsilon_2)}{d\epsilon_2} = \epsilon_2 e^{-\epsilon_2} e^{-e^{-\epsilon_2}}$$

Given $f(\epsilon_1, \epsilon_2) = f_1(\epsilon_1) f_2(\epsilon_2)$, we conclude that $\epsilon_1 \perp\!\!\!\perp \epsilon_2$.

Question (b)

$$
\begin{aligned}
Pr\Big(\text{i chooses 0}\Big) &= Pr\Big(U_{i0} \geq U_{i1} \text{ and } U_{i0} \geq U_{i2}\Big) \\
&= Pr\Big(\delta_0 + \epsilon_0 \geq \delta_1 + \epsilon_{i1} \text{ and } \delta_0 + \epsilon_0 \geq \delta_2 + \epsilon_{i2}\Big) \\
&= Pr\Big(\epsilon_{i1} \leq \delta_0 - \delta_1 + \epsilon_0 \text{ and } \epsilon_{i2} \leq \delta_0 - \delta_2 + \epsilon_0\Big) \\
&= \int F\Big(\delta_0 - \delta_1 + \epsilon_0, \delta_0 - \delta_2 + \epsilon_0)dF(\epsilon_0\Big) \\
&= \int_{-\infty}^{\infty} exp\Big\{ - \Big[exp\Big(- \frac{\delta_0 - \delta_1 + \epsilon_0}{\rho}\Big) + exp\Big(- \frac{\delta_0 - \delta_2 + \epsilon_0}{\rho}\Big)\Big]^{\rho}\Big\} dexp[-exp(\epsilon_0)] \\
&= \int_{-\infty}^{\infty} exp\Big\{ - exp(\epsilon_0)\Big[exp(- \frac{\delta_0 - \delta_1}{\rho}) + exp(- \frac{\delta_0 - \delta_2}{\rho})\Big]^{\rho}\Big\} dexp[-exp(\epsilon_0)] \\
&= \frac{1}{\Big[exp(- \frac{\delta_0 - \delta_1}{\rho}) + exp(- \frac{\delta_0 - \delta_2}{\rho})\Big]^{\rho} + 1} \\
&= \frac{1}{exp(-\delta_0)\Big[exp(\frac{\delta_1}{\rho}) + exp(\frac{\delta_2}{\rho})\Big]^{\rho} + 1} \\
&= \frac{exp(\delta_0)}{\Big[exp(\frac{\delta_1}{\rho}) + exp(\frac{\delta_2}{\rho})\Big]^{\rho} + exp(\delta_0)}
\end{aligned}
$$

Question (c)

Given that "1" (red bus) and "2" (blue bus) are in the same nest, and that "0" (car) is in another nest, we can derive the within-nest choice probability as

$$
\begin{aligned}
&Pr\Big(\text{i chooses 1} \Big| \text{ i does not chooses 0}\Big) \\
&= Pr\Big(U_{i1} \geq U_{i2}\Big) \\
&= Pr\Big(\delta_1 + \epsilon_{i1} \geq \delta_0 + \epsilon_{i0}\Big) \\
&= \int_{-\infty}^{\infty} F_1\Big(\epsilon, \delta_1 - \delta_2 + \epsilon\Big) d\epsilon
\end{aligned}
\tag{1}
$$

where $F_1\Big(\epsilon_1, \epsilon_2\Big) = \frac{\partial F(\epsilon_1, \epsilon_2)}{\partial \epsilon_1} = F(\epsilon_1, \epsilon_2)\Big(exp(- \frac{\epsilon_1}{\rho}) + exp(- \frac{\epsilon_1}{\rho})\Big)^{\rho - 1} exp\Big(- \frac{\epsilon_1}{\rho}\Big)$

Then denote $K = exp\big(-\frac{\delta_1-\delta_2}{\rho}\big)+1$ and $t = exp\big(-\frac{\epsilon}{\rho}\big)$, we can simplify the integral as follows

$$\begin{aligned}
&Pr\Bigg(\text{i chooses 1}\;\Bigg|\;\text{i does not chooses 0}\Bigg)\\
&=\frac{\rho}{K}\int_{t=0}^{+\infty} exp\Bigg(-(tK)^{\rho}\Bigg)(tK)^{\rho-1}d(tK)\\
&=-\frac{1}{K}\int_{t=0}^{+\infty} exp\Bigg(-(tK)^{\rho}\Bigg)d\Bigg(-(tK)^{\rho}\Bigg)\\
&=\frac{1}{K}\\
&=\frac{1}{1+exp\big(-\frac{\delta_1-\delta_2}{\rho}\big)}\\
&=\frac{exp\big(\frac{\delta_1}{\rho}\big)}{exp\big(\frac{\delta_1}{\rho}\big)+exp\big(\frac{\delta_2}{\rho}\big)}
\end{aligned} \tag{2}$$

2 Price Elasticities in Logit Demand Model

In the system of homogeneous consumers, the Marshallian demand is proportionate to the market share s_j, which is approximate to the individual choice probability $Pr(i \to j)$. Given the consideration set $\mathcal{J}_t$, the regularity condition holds that for any alternative $j \in \mathcal{J}_t$, the price after infinitesimal changes is still smaller or equal to the income, $p_j + \Delta p \leq m$.

$$\frac{\partial Pr(i \to j)}{\partial p_c}\frac{p_c}{s_j} = \begin{cases} -\alpha p_j(1-s_j) & \text{if } j = c \\ \alpha p_c s_c & \text{if } j \neq c \end{cases}$$

This is the legitimate Marshallian own and cross price elasticities by using choice probability as the infinitesimal changes in prices p_c leads to changes in the individual choice probabilities $Pr(i \to j)$, which then translates to proportional changes in the Marshallian demand.

However, this is not the legitimate Hicksian own and cross price elasticities. Hicksian demand function or compensated demand function for a good is his quantity demanded as part of the solution to minimizing his expenditure on all goods while delivering **a fixed level of utility**. This contradicts with the underlining assumption of the logit demand function.

$$Pr\Bigg(i \to j\Bigg) = Pr\Bigg(U_{ij} \geq U_{ik}\forall k \neq j \in \mathcal{J}_t\Bigg)$$

As a consequence, there is no utility level $\bar{U}$ which could serve as the fixed threshold in the sense that

$Pr\Big(i \to j\Big) = Pr\Big(U_{ij} \geq \bar{U}\Big)$. In other words, $U_{ij} \geq \bar{U}$ cannot rule out the possibility that $U_{ik} > U_{ij} \geq \bar{U}$. In summary, the proposed price elasticities are illegitimate Hiscksian own and cross elasticities as only the **relative** utilities matter in the logit demand models, rather than the **absolute** levels.

3 Linear Probability Model

Question (a)

$$\begin{aligned} P_i &= Pr\Big\{y_i = x_i'\beta + u_i > 0\Big|x_i\Big\} \\ &= Pr\Big\{u_i > -x_i'\beta\Big|x_i\Big\} \end{aligned} \tag{3}$$

Given $u_i \sim^{iid} Uniform[-0.5, 0.5]$, we have

$$\begin{aligned} P_i &= Pr\Big\{u_i > -x_i'\beta\Big|x_i\Big\} \\ &= 0.5 + x_i'\beta \end{aligned} \tag{4}$$

The error term in the linear probability model is

$$e_i = d_i - P_i = \begin{cases} 0.5 - x_i'\beta & \text{if } y_i = x_i'\beta + u_i > 0 \\ -0.5 - x_i'\beta & \text{otherwise} \end{cases}$$

Then we derive the conditional expectation of the error terms as

$$\begin{aligned} \boldsymbol{E}\Big(e_i\Big|x_i\Big) &= Pr\Big\{d_i = 1\Big\}\boldsymbol{E}\Big(e_i\Big|x_i, d_i = 1\Big) + Pr\Big\{d_i = 0\Big\}\boldsymbol{E}\Big(e_i\Big|x_i, d_i = 0\Big) \\ &= \Big(0.5 - x_i'\beta\Big)\Big(0.5 + x_i'\beta\Big) + \Big(-0.5 - x_i'\beta\Big)\Big(0.5 - x_i'\beta\Big) \\ &= 0 \end{aligned}$$

Question (b)

Then we derive the conditional variance of the error terms as

$$
\begin{aligned}
\sigma_i^2 = \boldsymbol{Var}\left(e_i \middle| x_i\right) &= \boldsymbol{E}\left(e_i^2 \middle| x_i\right) + \left\{\boldsymbol{E}\left(e_i \middle| x_i\right)\right\}^2 \\
&= \boldsymbol{E}\left(e_i^2 \middle| x_i\right) \\
&= Pr\left\{d_i = 1\right\}\boldsymbol{E}\left(e_i^2 \middle| x_i, d_i = 1\right) + Pr\left\{d_i = 0\right\}\boldsymbol{E}\left(e_i^2 \middle| x_i, d_i = 0\right) \\
&= \left(0.5 - x_i'\beta\right)^2\left(0.5 + x_i'\beta\right) + \left(-0.5 - x_i'\beta\right)^2\left(0.5 - x_i'\beta\right) \\
&= 0.25 - \left(x_i'\beta\right)^2
\end{aligned}
\tag{5}
$$

Question (c)

$$
\begin{aligned}
\boldsymbol{Cov}\left(e_i, e_j \middle| x_i, x_j\right) &= \boldsymbol{E}\left(e_i, e_j \middle| x_i, x_j\right) + \left\{\boldsymbol{E}\left(e_i \middle| x_i\right)\boldsymbol{E}\left(e_j \middle| x_j\right)\right\} \\
&= \boldsymbol{E}\left(e_i, e_j \middle| x_i, x_j\right) \\
&= Pr\left\{d_i = 1, d_j = 1\right\}\boldsymbol{E}\left(e_i e_j \middle| x_i, x_j, d_i = 1, d_j = 1\right) \\
&+ Pr\left\{d_i = 1, d_j = 0\right\}\boldsymbol{E}\left(e_i e_j \middle| x_i, x_j, d_i = 1, d_j = 0\right) \\
&+ Pr\left\{d_i = 0, d_j = 1\right\}\boldsymbol{E}\left(e_i e_j \middle| x_i, x_j, d_i = 0, d_j = 1\right) \\
&+ Pr\left\{d_i = 1, d_j = 1\right\}\boldsymbol{E}\left(e_i e_j \middle| x_i, x_j, d_i = 0, d_j = 0\right) \\
&= 0
\end{aligned}
\tag{6}
$$

Combining (5)(6), we know that the error terms are heterogeneous and uncorrelated.

$$
\boldsymbol{\Omega} = Diag\left(\sigma_i^2\right) \tag{7}
$$

The regression specification of the linear probability model is as follows

$$
d_i = P_i + e_i = 0.5 + x_i'\beta + e_i
$$

The Ordinary Least-Square estimator

$$\hat{\beta}_{OLS} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'(\boldsymbol{d} - \mathbf{0.5})\right)$$

The estimated asymptotic variance matrix of $\hat{\beta}_{OLS}$ with heteroskedasticity robust standard errors

$$\hat{V}[\hat{\beta}_{OLS}] = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\hat{\boldsymbol{\Omega}}\boldsymbol{X}\right)\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \tag{8}$$

where $\hat{\boldsymbol{\Omega}} = Diag(\hat{e}_i^2)$ and $\hat{e}_i = d_i - 0.5 - x_i'\hat{\beta}_{OLS}$.

Given $e_i\Big|x_i \rightarrow^d N\left(0, \boldsymbol{\Omega}\right)$ and $\boldsymbol{\Omega} = Diag\left(\sigma_i^2\right)$, we have

$$plim\hat{V}[\hat{\beta}_{OLS}] = plimV[\hat{\beta}_{OLS}] = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X}\right)\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

Question (d)

Given that the errors are heterogeneous and uncorrelated, we specify $V[e\big|x] = exp(x'\gamma)$.

Then use the non-linear least-square regression of $\hat{e}_i^2 = (d_i - 0.5 - x_i'\hat{\beta})^2$ on $exp(x'\gamma)$ to derive $\hat{\gamma}$, which is the consistent estimator of γ. The estimator error matrix is

$$\hat{\boldsymbol{\Omega}} = \hat{\boldsymbol{\Omega}}(\hat{\gamma}) = exp(x'\hat{\gamma})$$

We have the feasible generalized least-square estimator $\hat{\beta}_{FGLS}$, which is more efficient than the given $\hat{\beta}$ as

$$\hat{\beta}_{FGLS} = \left(\boldsymbol{X}'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\hat{\boldsymbol{\Omega}}^{-1}(\boldsymbol{d} - \mathbf{0.5})\right)$$

The estimated asymptotic variance of $\hat{\beta}_{FGLS}$ is

$$\hat{V}[\hat{\beta}_{FGLS}] = \left(\boldsymbol{X}'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{X}\right)^{-1} \tag{9}$$

Question (e)

Given that $\mathbf{\Omega}(\gamma)$ is correctly specified and that $\hat{g}amma$ is consistent for γ, we have (Cameron & Trivedi, page 82)

$$\sqrt{N}\left(\hat{\beta}_{FGLS}-\beta\right)\to^{d} N\left[0,\left(plimN^{-1}\boldsymbol{X}'\mathbf{\Omega}^{-1}\boldsymbol{X}\right)^{-1}\right]$$

This implies that

- $\hat{\beta}_{FGLS}$ is a consistent estimator of β. $\hat{\beta}_{FGLS}\to_{p}\beta$ as $n\to\infty$
- $\hat{\beta}_{FGLS}$ is second-moment efficient as $\hat{\beta}_{FGLS}$ has the same limiting variance matrix as $\hat{\beta}_{GLS}$.

In summary, $\hat{\beta}_{FGLS}$ possesses the same asymptotic properties as $\hat{\beta}_{MLE}$.

4 Multinomial Logit and MLE Estimation

Question (a)

The likelihood function of the binary choice models for individual i is

$$L\left(\{y_{i,j},x_j\}_{j\in J}\middle|\beta\right)=\prod_{i=1}^{N}\prod_{j\in J}\left[\frac{exp(x'_j\beta)}{\sum_{k=0}^{J}exp(x'_k\beta)}\right]^{y_{i,j}}$$

The log-likelihood function of observing $\{y_{i,j},x_j\}_{j\in J}$ for individuals $i=1,2,...,N$ is

$$lnL\left(\{y_{i,j},x_j\}_{j\in J}\middle|\beta\right)=\sum_{i=1}^{N}\sum_{j\in J}\left\{y_{i,j}ln(\frac{exp(x'_j\beta)}{\sum_{k\in J}exp(x'_k\beta)})\right\}$$

Question (b)

Denote $G(x'_j\beta)=\frac{exp(x'_j\beta)}{\sum_{k\in J}exp(x'_k\beta)}$ and

$$\begin{aligned}g(x'_j\beta)&=\frac{dG(x'_j\beta)}{d\beta}\\&=\frac{exp(x'_j\beta)\sum_{k\in J}\left[(x_j-x_k)exp(x'_k\beta)\right]}{\left(\sum_{k\in J}exp(x'_k\beta)\right)^2}\end{aligned}$$

Then we can rewrite $lnL\left(\{y_{i,j}, x_j\}_{j\in J}\middle|\beta\right)$ as

$$lnL\left(\{y_{i,j}, x_j\}_{j\in J}\middle|\beta\right) = \sum_{i=1}^{N}\sum_{j\in J}\left\{y_{i,j}ln\left(G(x_j'\beta)\right)\right\}$$

Taking derivatives of $lnL\left(\{y_{i,j}, x_j\}_{j=0}^{J}\middle|\beta\right)$ with respects to β, we can get the score function of the likelihood of observing the data $\{y_{i,j}, x_j\}_{j\in J}$ for individuals $i = 1, 2, ..., N$ as follows.

$$\begin{aligned} s\left(\beta\middle|\{y_{i,j}, x_j\}_{j\in J}\right) &= \frac{dlnL\left(\{y_{i,j}, x_j\}_{j\in J}\middle|\beta\right)}{d\beta} \\ &= \sum_{i=1}^{N}\sum_{j\in J}\left\{y_{i,j}\frac{g(x_j'\beta)}{G(x_j'\beta)}\right\} \end{aligned}$$

Question (c)

The information matrix of the (unconditional) likelihood of observing the data for individuals $i = 1, 2, ..., N$

$$I(\beta) = E_\beta\left(s\left(\beta\middle|\{y_{i,j}, x_j\}_{j\in J}\right)s\left(\beta\middle|\{y_{i,j}, x_j\}_{j\in J}\right)'\right) = \sum_{i=1}^{N}\sum_{j\in J}E_\beta\left\{y_{i,j}\frac{g(x_j'\beta)g(x_j'\beta)'}{G^2(x_j'\beta)}\right\}$$

Under mild regularity conditions,

$$I(\beta) = -E_\beta\left(\frac{\partial}{\partial\beta}s\left(\beta\middle|\{y_{i,j}, x_j\}_{j\in J}\right)\right) = -E_\beta\left(\frac{\partial^2}{\partial\beta^2}lnL\left(\{y_{i,j}, x_j\}_{j\in J}\middle|\beta\right)\right)$$