# 0 0 ECON0021 Microeconometrics Index Splited W56789

## Table of Content

## Table of Contents

---

# Week 5: MLE & QMLE

## *5 MLE & QMLE*

### *Linear Probability Model*

- Linear Probability Model:

$$Pr(Y=1|X) = \mathbb{E}[Y|X] = X'\beta \iff Y = X'\beta + u, \mathbb{E}[u|X] = 0$$

- Problems with Linear Probability Model
    - Fitted Pr outside $[1,0]$: $\mathbb{E}[Y_i|X_i] = X_i'\beta \notin [0,1]$
    - Constant marginal effects (unrealistic)
    - Heteroskedasticity

### *MLE*

- Maximum Likelihood
    - Parametric: the population distribution belongs to a family of distributions fully characterised by a finite number of unknown parameters $\theta \in \mathbb{R}^p$
        - i.e. if we know a finite number of parameters specify the distribution, we know the full distribution
    - We want to estimate $\theta_0$, which is the true vector of parameters in the DGP
    - Given a sample, MLE Estimator is defined by the parameter value that maximises the likelihood (the probability of observing the sample)

#### *General (conditional) MLE Steps - My Personal Workflow*

1. Specify a parametric distribution $f_z(z;\theta)$
2. Specify a functional form for conditional likelihood $f_{Y|X}(y|x;\theta)$
3. Calculate individual log likelihood (pdf/pmf)

$$\log f_{Y|X}(y_i|x_i;\theta)$$

4. Log likelihood function for iid data:

$$\log \mathcal{L}_n(y_i|x_i;\theta) = \sum_{i=1}^{n} \log f_{Y|X}(y_i|x_i;\theta)$$

5. Calculate the score from *individual* likelihood

$$s(y_i, x_i; \theta) = \frac{\partial \log f_{Y|X}(y_i|x_i, \theta)}{\partial \theta}$$

caveat: differentiate to the *exact* parameters of the distribution (e.g. $\sigma^2$ instead of $\sigma$ for a Normal distribution)

6. Obtain the MLE estimator $\hat{\theta}$ from FOC: set expectation of score to be 0

$$\frac{\partial}{\partial \theta}\mathbb{E}[\log f_{Y|X}(y_i|x_i, \theta_0)] = \mathbb{E}[s(y_i|x_i; \theta_0)] = 0$$
$$\implies 0 = \frac{\partial}{\partial \theta}\left\{\frac{1}{n}\sum_{o=1}^{n} \log f_{Y|X}(y_i|x_i; \hat{\theta}_{MLE})\right\} = \frac{1}{n}\sum_{i=1}^{n} s(y_i|x_i; \hat{\theta}_{MLE}) = 0$$

(alternatively: use FOC: $\frac{\partial \log \mathcal{L}_n(Y_i|x_i;\theta)}{\partial \theta} = 0$)

7. Calculate Hessian from score

$$H(y_i, x_i; \theta) = \frac{\partial s(y_i, x_i; \theta)}{\partial \theta'}$$

8. Calculate Fisher Information Matrix from Hessian using the Information Matrix Equality (not valid in QMLE)

$$\mathcal{I}(\theta) = -\mathbb{E}_\theta[H(y_i, x_i; \theta)]$$

9. Calculate asymptotic variance from Hessian (use true parameter $\theta_0$ here!)

$$Avar\left(\sqrt{n}\left(\hat{\theta}_{MLE} - \theta_0\right)\right) = \mathcal{I}(\theta_0)^{-1}; \widehat{Avar}\left(\hat{\theta}_{MLE} - \theta_0\right) = \frac{\mathcal{I}(\hat{\theta}_0)^{-1}}{n}$$

10. Obtain asymptotic distribution

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta) \to^d \mathcal{N}(0, \mathcal{I}(\theta_0)^{-1})$$

*General MLE Steps:*

1. Specify a parametric distribution $f_z(z; \theta)$
2. Calculate the joint likelihood function (sample likelihood)

$$\mathcal{L}_n(z; \theta) = \prod_{i=1}^{n} f_z(z_i, \theta)$$

another way is to calculate the log of individual likelihood, and then sum up

3. Take log (monotonic transformation)

$$log\mathcal{L}_n(z; \theta) = \sum_{i=1}^{n} logf_z(z_i, \theta)$$

4. Maximise with respect to $\theta$ (often take FOCs)

$$\hat{\theta}_{MLE} = arg\max_{\theta\in\Theta} log\mathcal{L}_n(z,\theta)$$

caveat: differentiate to the *exact* parameters of the distribution (e.g. $\sigma^2$ instead of $\sigma$ for a Normal distribution)

5. Calculate the Fisher Information Matrix $\rightarrow$ Asymptotic distribution of MLE using following step below

### *Consistency*

- Consistency
  - Identification (Requirement): different parameters generate different distributions:

$$\theta \neq \theta' \implies f_z(z,\theta) \neq f_z(z,\theta')$$

  - If identification hold, $\theta_0 = arg\max_{\theta\in\Theta}\mathbb{E}[\log f_Z(z_i,\theta))$ uniquely maximises $\mathbb{E}[logf_z(z_i;\theta)]$ (a mild additional requirement: $\mathbb{E}[logf_z(z_i;\theta)] < \infty$) and its sample analogue:

$$\hat{\theta}_{MLE} = arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} logf_z(z_i,\theta) \rightarrow^p arg\max_{\theta\in\Theta}\mathbb{E}[logf_z(z_i,\theta)]$$

    i.e. $\hat{\theta}_{MLE} \rightarrow^p \theta_0$

### *Score*

- Score
  - Score

$$s(z_i,\theta) = \frac{\partial logf_z(z_i,\theta)}{\partial\theta} = \begin{bmatrix} \frac{\partial logf_z(z_i,\theta)}{\partial\theta_1} \\ \cdots \\ \frac{\partial logf_z(z_i,\theta)}{\partial\theta_p} \end{bmatrix}_{p\times 1}$$

  - FOC implies that (for population)

$$0 = \frac{\partial}{\partial\theta}\mathbb{E}[logf_z(z_i,\theta_0)] = \mathbb{E}[s(z_i,\theta_0)]$$

    for sample:

$$0 = \frac{\partial}{\partial\theta}\left\{\frac{1}{n}\sum_{i=1}^{n} logf_z(z_i,\hat{\theta}_{MLE})\right\} = \frac{1}{n}\sum_{i=1}^{n} s(z_i,\hat{\theta}_{MLE})$$

  - If the samples perfectly fit the estimated distribution, then each score $= 0$

### *Fisher Information and AVAR*

- Fisher Information and Asymptotic Variance

- Fisher Information Matrix is the expectation of the outer product of the score vector:

$$\mathcal{I}_{p\times p}(\theta) = \mathbb{E}_\theta[s(z_i,\theta)s(z_i,\theta)'] = Var_\theta[s(z_i,\theta)] = \int s(z_i,\theta)s(z_i,\theta)'f_z(z_i,\theta)dz_i$$

- Asymptotic variance of MLE is the inverse of the information matrix evaluated at the *true* $\theta_0$:

$$Avar\left(\sqrt{n}\left(\hat{\theta}_{MLE} - \theta_0\right)\right) \rightarrow^p \mathcal{I}(\theta_0)^{-1}$$

and the Asymptotic Distribution of MLE is

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta_0) \rightarrow^d \mathcal{N}(0, \mathcal{I}(\theta_0)^{-1})$$

### *Hessian and Information Matrix Equality*

- Hessian and Information Matrix Equality
    - Hessian

$$H(z_i,\theta) = \frac{\partial^2 log f_z(z_i,\theta)}{\partial\theta\partial\theta'} = \begin{bmatrix} \frac{\partial^2 log f_z(z_i,\theta)}{\partial\theta_1^2} & \dots & \frac{\partial^2 log f_z(z_i,\theta)}{\partial\theta_1\partial\theta_p} \\ \dots & \dots & \dots \\ \frac{\partial^2 log f_z(z_i,\theta)}{\partial\theta_p\partial\theta_1} & \dots & \frac{\partial^2 log f_z(z_i,\theta)}{\partial\theta_p^2} \end{bmatrix}$$

    - and note that Hessian is the derivative of Score:

$$H(z_i,\theta) = \frac{\partial}{\partial\theta'} s(z_i,\theta)$$

    - Information Matrix Equality: Fisher Information Matrix = -Expectation of Hessian

$$\mathcal{I}(\theta) = -\mathbb{E}_\theta[H(z_i,\theta)] = -\int H(z_i,\theta)f_z(z_i,\theta)dz_i$$

- Inference
    - $$\sqrt{n}\left(\hat{\theta}_{MLE} - \theta_0\right) \rightarrow^d \mathcal{N}\left(0, \mathcal{I}(\theta_0)^{-1}\right)$$
    - We can estimate $\mathcal{I}(\theta_0)$ by:
        1. $\mathcal{I}(\hat{\theta}) = -\frac{1}{n}\sum_{i=1}^n H(z_i,\hat{\theta}_{MLE})$ (Information Matrix Equality)
        2. or $\mathcal{I}(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^n s(z_i,\hat{\theta}_{MLE})s(z_i,\hat{\theta}_{MLE})'$
    - and calculate $\widehat{Avar}(\hat{\theta}_{MLE}) = \hat{I}(\theta_0)^{-1}$

### *For Scalars*

- Score: $s(z_i,\theta) = \frac{\partial log f_z(z_i,\theta)}{\partial\theta}$
- Hessian: $H(z_i,\theta) = \frac{\partial^2 log f_z(z_i,\theta)}{\partial\theta^2} = \frac{\partial s(z_i,\theta)}{\partial\theta}$

- Fisher Information Matrix: $\mathcal{I}(\theta) = \underbrace{Var_\theta[s(z_i,\theta)]}_{=E_\theta[(s(z_i,\theta))^2]} \underbrace{= -\mathbb{E}_\theta[H(z_i,\theta)]}_{\text{Information Matrix Equality}}$
- Asymptotic Variance: $Avar\left(\sqrt{n}\left(\hat{\theta}_{MLE} - \theta_0\right)\right) = \mathcal{I}(\theta_0)^{-1}$

## *Quasi MLE*

- Quasi Maximum Likelihood Estimator (QMLE)
- $$\sqrt{n}\left(\hat{\theta}_{MLE} - \theta_0\right) \to^d \mathcal{N}\left(0, \Sigma\right)$$

  where

  $$\Sigma = \{\mathbb{E}[H(z_i,\theta_0)]\}^{-1} Var[s(z_i,\theta_0)]\{\mathbb{E}[H(z_i,\theta_0)]\}^{-1}$$

  is the Misspecification Robust SE with its sample analogue:

  $$\hat{\Sigma} = \left[\frac{1}{n}\sum_{i=1}^{n} H(z_i,\hat{\theta})\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n} s(z_i,\hat{\theta})s(z_i,\hat{\theta})'\right]\left[\frac{1}{n}\sum_{i=1}^{n} H(z_i,\hat{\theta})\right]^{-1}$$

- The *information matrix equality does not hold for QMLE*, so it can be used as a test for misspecification: information matrix test:

  $$\text{Correct Specification} \implies -\frac{1}{n}\sum_{i=1}^{N} H(w_i,\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{N} s(w_i,\hat{\theta})s'(w_i,\hat{\theta})'$$

## *Conditional MLE*

- Conditional MLE: often $z_i = (y_i, x_i')'$ and we only specify $f_{Y|X}(y_i|x_i;\theta)$ (allows for different distributions for different $x_i$)
- In reality, we make the following assumptions:
  1. Distribution Assumption: specify a distribution $f_z(z_i,\theta)$
  2. Functional Form Assumption: specify $f_{Y|X}(f_i|x_i,\theta)$
- Relationship between con/uncon liklihood/pdf:

  $$\underbrace{L_N(x_i,y_i,\theta)}_{Uncon.\ Likelihood} = \prod_{i=1}^{N}\underbrace{f_{XY}(y_i,x_i,\theta)}_{Uncon.\ PDF} = \prod_{i=1}^{N}\underbrace{f_{Y|X}(y_i|x_i,\theta)}_{Cond.PDF} \times \underbrace{f_X(x_i)}_{Marginal\ Dist.\ of\ x}$$

- Conditional MLE maximises conditional log-likelihood (because $\log f_Z(z_i;\theta) = \log f_{Y|X}(y_i|x_i;\theta) + \log f_X(x_i)$ and $\log f_X(x_i)$ does not depend on $\theta$):

  $$\hat{\theta}_{MLE} = arg\max_{\theta\in\Theta}\sum_{i=1}^{n} log f_{Y|X}(f_i|x_i,\theta)$$

- Chain Rule for Vectors

$$\frac{\partial}{\partial \beta} F(x^\intercal \beta) = x F'(x^\intercal \beta); \frac{\partial}{\partial \beta^T} F(x^\intercal \beta) = x^\intercal F'(x^\intercal \beta)$$

---

# Week 6: Binary Choice Model

## 6 Binary Choice Model - A

### Binary Choice Targets of Interest

- Binary Choice
- We want to learn:
  - $Pr(Y=1|X=x) = \mathbb{E}[Y|X=x]$: the conditional probability given an individual's characteristics
  - $\frac{\partial Pr(Y=1|X=x)}{\partial x_l}$: the partial marginal effect of changes in $x_l$

### Linear Probability Model

- Linear Probability Model:

$$y_i = x_i'\beta + \epsilon_i, \mathbb{E}[\epsilon_i|x_i] = 0 \quad \text{or} \quad \mathbb{E}[y_i|x_i] = x_i'\beta$$

- Note that because $y_i$ is binomial (conditional on $x_i$), so:

$$\begin{cases} x_i'\beta = Pr(y_i = 1|x_i) \\ Var[\epsilon_i|x_i] = Pr(y_i = 1|x_i)(1 - Pr(y_i = 1|x_i)) \end{cases}$$

- Problems: Heteroskedasticity, Out-of-bound Predictions, Constant Marginal Effects

### Reduced Form Approach

- Reduced Form Approach
- Use some increasing functions $\mathcal{F} : \mathbb{R} \to [0,1]$ to model:

$$Pr(y_i = 1|x_i) = \mathcal{F}(x_i'\beta)$$

- A natural choice is the CDF: $F(z) = Pr(\epsilon \leq z)$ of a random variable $\epsilon$, and the Conditional Likelihood takes the form of:

$$Pr(y_i|x_i, \beta) = F(x_i'\beta)^{y_i}[1 - F(x_i'\beta)]^{1-y_i}$$

- ↓ adds economics to such approach

### Structural Threshold Crossing Models

#### *Threshold Crossing Model*

- Threshold Crossing Model: Suppose

$$y_i^* = x_i'\beta - \epsilon_i, \quad y_i = \begin{cases} 1 & \text{if } y_i^* \geq 0 \\ 0 & \text{if } y_i^* < 0 \end{cases}$$

  then

$$Pr(y_i = 1|x_i) = F(x_i'\beta)$$

  where $F$ is the CDF for $\epsilon$
- Conditional likelihood takes the form of:

$$Pr(y_i|x_i, \beta) = F(x_i'\beta)^{y_i}\Big[1 - F(x_i\beta)\Big]^{1-y_i}$$

- This can be motivated by the Random Utility Model (an example of threshold crossing model):

$$Pr(y_i = 1|x_i) = Pr(\underbrace{U_{i1} - U_{i0}}_{y^*} \geq 0|x_i)$$

  where $U_{i1}$ and $U_{i0}$ are utilities for $y_i = 1$ and $y_i = 0$

### *Specifications*

- Specifications
- Probit Model: $\epsilon \sim \mathcal{N}(0, 1)$ with CDF $F(z) = \Phi(z)$:

$$Pr(y_i = 1|x_i) = \Phi(x_i'\beta) \text{ where } \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz$$

- Logit Model: $\epsilon \sim$ Logistic distribution with CDF $F(z) = \Lambda(z)$:

$$Pr(y_i = 1|x_i) = \Lambda(x_i'\beta) = \frac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)} \text{ where } \Lambda(z) = \frac{\exp(z)}{1 + \exp(z)}$$

  - Logistic distribution is similar to Normal distribution except with $Var = \frac{\pi^3}{3}$

### *Identification*

- Identification: we must impose location and scale normalisations to fit the logit/probit
- Our actual model could be:

$$y_i^* = \beta_0 + x_i'\beta - \epsilon_i, \quad y_i = \begin{cases} 1 & \text{if } y_i^* \geq 0 \\ 0 & \text{if } y_i^* < 0 \end{cases}$$

  where $E[\epsilon_i] = \mu, Var[\epsilon_i] = \sigma^2$ and its distribution is specified
- Location normalisation: adding the same constant to $\beta_0$ and to $\epsilon_i$ will not change $y_i$, so there is no generality to assume $\mu = 0$
- Scale normalisation: multiplying $(\beta_0, \beta')$ and $\epsilon_i$ by the same positive constant will not change $y_i$, so there is no loss of generality to assume $\sigma = 1$

- Therefore, $\beta_0, \beta$ and $\epsilon_i$ are not uniquely identified, and it's meaningless to look/compare at their corresponding estimates directly. (Note that the sign does have an immediate interpretation)

## *Maximum Likelihood*

- Maximum Likelihood
- Likelihood:

$$L_n(\beta) = \prod_{i=1}^{n} F(x_i'\beta)^{y_i}[1 - F(x_i'\beta)]^{1-y_i}$$

- Log-likelihood:

$$\log L_n(\beta) = \sum_{i=1}^{n} y_i \log F(x_i'\beta) + (1 - y_i) \log[1 - F(x_i'\beta)]$$

- FOC:

$$0 = \frac{\partial \log L_n(\hat{\beta})}{\partial \beta} = \sum_{i=1}^{n} \frac{F'(x_i'\hat{\beta})}{F(x_i'\beta)(1 - F(x_i'\beta))} \Big(y_i - F(x_i'\beta)\Big) x_i$$

- Logit FOC:

$$0 = \sum_{i=1}^{n} \Big(y_i - \Lambda(x_i'\hat{\beta})\Big) x_i$$

## *Marginal Effects*

- Marginal Effects
  - Conditional Probability

$$Pr(y_i = 1|x_i) = \begin{cases} x_i'\beta & \text{, for LPM} \\ \underbrace{F(x_i'\beta_l)}_{cdf} = \begin{cases} \Phi(x_i'\beta) & \text{, for Probit} \\ \Lambda(x_i'\beta) = \frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)} & \text{, for Logit} \end{cases} \end{cases}$$

  - Marginal Effects for continuous $x_l$:

$$\frac{\partial Pr(y_i = 1|x_i)}{\partial x_{il}} = \begin{cases} \beta_l & \text{, for LPM} \\ \underbrace{f(x_i'\beta)}_{pdf}\beta_l = \begin{cases} \phi(x_i'\beta)\beta_l & \text{, for Probit} \\ \frac{\exp(x_i'\beta)}{[1+\exp(x_i'\beta)]^2}\beta_l & \text{, for Logit} \end{cases} \end{cases}$$

  - Marginal Effect for Dummy $x_l$:

$$Pr(y_i = 1|x_{il} = 1, x_{i,others}) - Pr(y_i = 1|x_{il} = 0, x_{i,others})$$

- Average Marginal Effects

- Since marginal effects are heterogenous, and magnitude of $\beta_l$ has no immediate interpretation (because of Identification Issue), we often report average marginal effects:

$$\mathbb{E}\left[\frac{\partial Pr(y_i = 1|x_i)}{\partial x_{il}}\right] = \mathbb{E}[f(x_i'\beta)\beta_l]$$

with its sample analogue:

$$\frac{1}{n}\sum_{i=1}^{n} f(x_i\hat{\beta})\hat{\beta}_l$$

---

# Week 7: Random Utility Model

## *7 Random Utility Model. Welfare Analysis. Binary Choice and Endogeneity - A*

### *Random Utility Model*

#### *Latent Utility Formulation*

- Random Utility Hypothesis: An individual has random utility that is drawn from a certain distribution each time a decision is made, and the individual makes a choice of $Y = 0$ or $Y = 1$ depending on which option provides a higher utility to her/him
- Latent Utility Formulation:

$$\begin{cases} U_1 = u_1(X) + \epsilon_1 \\ U_2 = u_0(X) + \epsilon_0 \end{cases}$$

where $u_1(X) = X'\gamma_1, u_0(X) = X'\gamma_0$ are non-random/deterministic latent utility dependent on observable characteristics $X$, and $\epsilon_1, \epsilon_0$ are random components of utility

- Then the choice probability of an individual with characteristic $X = x$ is:

$$\begin{aligned} Pr(Y = 1|X = x) &= Pr(U_1 \geq U_0|X = x) \\ &= Pr(x'\underbrace{(\gamma_1 - \gamma_0)}_{\beta} \geq \underbrace{\epsilon_0 - \epsilon_1}_{\Delta\epsilon}|X = x) \\ &= Pr(x'\beta \geq \Delta\epsilon|X = x) \end{aligned}$$

- assume $\Delta\epsilon \perp X$, this equals to $CDF_{\Delta\epsilon}(x'\beta)$
- with further distribution assumptions of $\Delta\epsilon$:

$$Pr(Y = 1|X = x) = \begin{cases} \Phi(x'\beta) & \text{, if } \epsilon_0, \epsilon_1 \sim \mathcal{N}\left(0, \frac{1}{2}\right) \rightarrow \Delta\epsilon \sim \mathcal{N}(0, 1) \\ \Lambda(x'\beta) & \text{, if } \epsilon_0, \epsilon_1 \sim \text{Type I extreme value} \rightarrow \epsilon \sim Logit \end{cases}$$

- Exogeneity: $\epsilon_1, \epsilon_2 \perp X$ (Statistical independence)

- Identification: parameters appear in both equations will be jointly identified (their differences are identified); *only parameters appear in only one of the equations will be uniquely identified*

### *Intra/Inter-Personal Random Utility*

- Intra- or Inter-personal random utility
- Intra-individual random utility: Even under a given choice situation and his/her characteristics, his/her choice of Y is not deterministic due to the random utility (random state of mind)
- Inter-individual random utility: Each individual has a nonrandom choice at each choice situation, while individuals with the same observed characteristics can make different choices due to heterogeneity in tastes.
- With cross-sectional data, those two are indistinguishable.

## *Welfare Analysis and Compensating Variation*

### *Framework*

- Welfare Analysis
- Consumer i choose between options $j \in \{0, 1\}$:

$$U_{ij} = u_j(I_i - p_{ij}, x_{ij}) + \epsilon_{ij}$$

  where $I_i$ is income, $p_{ij}$ is the price of option $j$, $x_{ij}$ are consumer and products' characteristics
- Before the intervention, the utility of consumer i is:

$$\max_{j \in \{0,1\}} \{u_j(I_i - p_{ij}, x_{ij}) + \epsilon_{ij}\}$$

- Intervention: new prices $p_{ij} \rightarrow \tilde{p}_{ij}$ and new characteristics $x_{ij} \rightarrow \tilde{x}_{ij}$
- Then Compensation Variation $CV_i(I_i, p_i, x_i, \epsilon_i)$ is the change in income needed to hold utility at its initial level:

$$\max_{j \in \{0,1\}} \{u_j(I_i - p_{ij}, x_{ij}) + \epsilon_{ij}\} = \max_{j \in \{0,1\}} \{u_j(I_i + CV_i - \tilde{p}_{ij}, \tilde{x}_{ij}) + \epsilon_{ij}\}$$

- Compensation variations are paid *ex-ante* (before choices are made)!

### *Estimation of Compensation Variation*

- Deal with heterogeneity in $\epsilon_i$: For every individual i, we compute:

$$\begin{aligned}\mathbb{E}[CV_i | I_i, p_i, x_i] &= \mathbb{E}[CV(I_i, p_i, x_i, \epsilon_i) | I_i, p_i, x_i] \\ &= \int CV(I_i, p_i, x_i, \epsilon_i) f_\epsilon(\epsilon)\, d\epsilon\end{aligned}$$

- This is often approximated by numerical integration (Monte-Carlo average):

$$\frac{1}{S}\sum_{s=1}^{S} CV(I_i, p_i, x_i, \epsilon_i^s)$$

where $\epsilon_e^s = (\epsilon_0^s, \epsilon_1^s)'$ are i.i.d. draws from the underlying distribution (e.g. probit $\mathcal{N}\left(0, \frac{1}{2}\right)$ or Type-1 extreme-value dist.) for a large number $S$
- With *known utility functions*, we *divide consumers according to their choice before/after intervention, and calculate the CV one by one*

- Deal with heterogeneity in individual characteristics $I_i, p_i, x_i$: LIE

$$\mathbb{E}[CV_i] = \mathbb{E}[\mathbb{E}[CV_i|I_i, p_i, x_i]]$$

- Thus, we can estimate $\mathbb{E}[CV_i]$ by:

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\underbrace{\frac{1}{S}\sum_{s=1}^{S} CV(I_i, p_i, x_i, \epsilon^S)}_{\hat{\mathbb{E}}[CV_i|I_i,p_i,x_i]}\right\}$$

### *Estimation Procedures in Action*

- Estimate all parameters by MLE
- Divide individuals into: always takers of each option and switchers
- Calculate the threshold of $\Delta\epsilon_i$ for each kind of individuals
- Calculate the $CV$ for each kind of individuals
    - For switchers, we need them to have the same utility with new options as before with old options
- Calculate with estimated model parameters:

$$\hat{\mathbb{E}}[CV_i|x_i] = \int CV(x_i) f_\epsilon(\epsilon)\, d\epsilon$$

- Calculate:

$$\mathbb{E}[CV_i] = \frac{1}{n}\sum_{i=1}^{N} \hat{\mathbb{E}}[CV_i|x_i]$$

## *Endogeneity in Binary Regressions*

### *Endogeneity*

- Binary regression with Endogeneity: individual observable characteristics $X$ and the unobservable $U$ are correlated:

$$Y = \mathbb{1}[\beta X + U \geq 0], X \not\perp U$$

- Example: $C$ is endogenous (selection on unobservable that affects $Y$)

$$Y = \mathbb{1}[\alpha C + \beta' W + U \geq 0], C \not\perp U$$

### *Bivariate Probit Model*

- Our model is:

$$\begin{cases} Y = \mathbb{1}[\alpha C + \beta' W + U \geq 0] \\ C = \mathbb{1}[\gamma' Z + V \geq 0] \end{cases}$$

where $dim(Z) > dim(W)$: $Z$ should contain at least 1 Instrumental Variable that is not included in $W$: it provides exogenous variations in regressor $C$ without causally affecting the outcome of interest $Y$ (The instrument has to satisfy:

1. Exclusion: $IV_i \perp (U_i, V_i)$
2. Relevance: $IV_i$ affects $C_i$

- Assume:

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

where $\rho \neq 0$ (otherwise $C$ will be exogenous)

- Likelihood of Bivariate Probit: $p(Y, C|W, Z; \theta)$ here $\theta = (\alpha, \beta, \gamma, \rho)'$
  - Conditional on $W, Z$, the realisation of $(Y, C)$ is determined by realisation of $(U, V)$

| $(Y, C)$ | Threshold | Likelihood |
|---|---|---|
| (0,0) | $U \leq -\beta' W, V \leq -\gamma' Z$ | $p_{00}(W, Z; \theta) = \int_{-\infty}^{-\beta' W} \int_{-\infty}^{-\gamma' Z} f_{U,V}(u, v; \rho)\, du\, dv$ |
| (1,0) | $U \geq -\beta' W, V \leq -\gamma' Z$ | $p_{10}(W, Z; \theta) = \int_{-\beta' W}^{+\infty} \int_{-\infty}^{-\gamma' Z} f_{U,V}(u, v; \rho)\, du\, dv$ |
| (0,1) | $U \leq -\alpha - \beta' W, V \geq -\gamma' Z$ | $p_{01}(W, Z; \theta) = \int_{-\infty}^{-\alpha-\beta' W} \int_{-\gamma' Z}^{+\infty} f_{U,V}(u, v; \rho)\, du\, dv$ |
| (1,1) | $U \geq -\alpha - \beta' W, V \geq -\gamma' Z$ | $p_{11}(W, Z; \theta) = \int_{-\alpha-\beta' W}^{+\infty} \int_{-\gamma' Z}^{+\infty} f_{U,V}(u, v; \rho)\, du\, dv$ |

- Joint Likelihood Function:

$$L_n(\theta) = \prod_{i=1}^{n} \{p_{11}(w_i, z_i; \theta)^{y_i c_i}\}\{p_{10}(w_i, z_i; \theta)^{y_i(1-c_i)}\}\{p_{01}(w_i, z_i; \theta)^{(1-y_i)c_i}\}\{p_{00}(w_i, z_i; \theta)^{(1-y_i)(1-c_i)}$$

and estimate $\theta = (\alpha, \beta, \gamma, \rho)'$ by MLE

- We can then estimate the Average Partial Effect:

$$\frac{1}{n} \sum_{i=1}^{n} \left(\hat{Pr}(Y = 1|C = 1, w_i) - \hat{Pr}(Y = 1|C = 0, w_i)\right)$$

# Week 8: Multinomial Choice (Vanilla Mn Logit & Too Flexible Mn Probit)

## 8 Multinomial Choice - A

### Notations

- $W_i$ are individual-specific characteristics
- $W_{ij}$ are individual- and choice-specific characteristics
- $X_i = (W_i, W_{i1}, \dots, W_{iJ})'$

### Multinomial Logit Model (Vanilla, Not Realistic)

#### Framework

- Multinomial Logit Model
- Latent Utility Formulation: The utility that individual i would get from choice j:

$$U_{ij} = V_j(X_i) + \epsilon_{ij} = V_{ij} + \epsilon_{ij}$$

- Assume: $\epsilon_{ij} \sim$ (*i.i.d. in i and j*) Type-I extreme value distribution:

$$\begin{aligned} Pr(Y_i = j | X_i = x) &= Pr(V_{ij} - V_{ik} \geq \epsilon_{ik} - \epsilon_{ij} \,\forall\, k | X_i = x) \\ &= \frac{\exp(V_{ij})}{\sum_{k=1}^{J} \exp(V_{ik})} \\ &= \frac{\exp(V_{ij} - V_{iJ})}{1 + \sum_{k=1}^{J-1} \exp(V_{ik} - V_{iJ})} \end{aligned}$$

  - 如果所有选项的$V$有common term，可以在第一步exp之后就约掉
- Parameterise $V_{ij} = \alpha_j + W_i'\beta_j + W_{ij}'\gamma$, then:

$$V_{ij} - V_{iJ} = (\alpha_j - \alpha_J) + W_i'(\beta_j - \beta_J) + (W_{ij} - W_{iJ})'\gamma$$

- Naturally, we normalise $\alpha_J = 0, \beta_J = 0$ and, for the outside option: $V_{iJ} = 0$, so:

$$V_{ij} - V_{iJ} = \alpha_j + W_i'\beta_j + W_{ij}\gamma$$

  - Note that we cannot incorporate products' characteristics that does not vary over individuals because they will be absorbed by $\alpha_j$

#### Estimation: MLE

- Estimation: MLE

- Likelihood function:

$$\begin{aligned} L_n(\theta) &= \prod_{i=1}^{n} Pr(Y_i = y_i | X_i = x_i; \theta) \\ &= \prod_{i=1}^{n} \frac{\exp(V_{y_i}(x_i;\theta))}{\sum_{k=1}^{J} \exp(V_k(x_i;\theta))} \end{aligned}$$

where $\theta = (\alpha_1, \ldots, \alpha_{J-1}, \beta_1, \ldots, \beta_{J-1}, \gamma')'$

### *Marginal Effects and Elasticities*

- Denote $P_j(x) = Pr(Y_i = j | X_i = x)$
- Marginal effects:

$$\frac{\partial P_j(x)}{\partial w_{kl}} = \begin{cases} \gamma_l P_j(x)(1 - P_j(x)) & , j = k \\ -\gamma_l P_j(x) P_k(x) & , \text{otherwise} \end{cases}$$

- Price elasticities ($w_{kl} = p_k, \gamma_l = -\alpha$):

$$\frac{\partial P_j(x)}{\partial p_k} \frac{p_k}{P_j(x)} = \begin{cases} -\alpha p_j(1 - P_j(x)) & , j = k \\ \alpha p_k P_k(x) & , \text{otherwise} \end{cases}$$

this shows the IIA property

### *Drawback: Substitution Patterns (IIA) and Own-Elasticities*

- Problem 1: Independence from Irrelevant Alternatives (IIA): for any two choices $j$ and $k$, the ratio of the choice probabilities:

$$\frac{Pr(Y_i = j|x)}{Pr(Y_i = k|x)} = \exp(V_{ij} - V_{ik})$$

*does not depend on the characteristics of other alternatives*

- Problem 2: Own-elasticities: most markets have many options, so $(1 - P_j(x))$ is close to a constant, so the *own-elasticity will be proportional to price*:

$$\frac{\partial P_j(x)}{\partial p_j} \frac{p_j}{P_j(x)} = -\alpha p_j \underbrace{(1 - P_j(x))}_{\text{close to constant}}$$

i.e. low price $\implies$ low elasticity (not realistic)

## *Flexible Substitution Pattern*

- Introduce correlations among the random components in the utility $\epsilon_{ij}$
- Methods: Multinomial Probit, Nested Logit, Random Coefficient Logit

### *Multinomial Probit (Too Flexible)*

- Very flexible

- Setup:

$$U_{ij} = V_{ij} + \epsilon_{ij}, \;\; \epsilon_i = \begin{pmatrix} \epsilon_{i1} \\ \dots \\ \epsilon_{iJ} \end{pmatrix} \sim N(0, \Sigma)$$

where $\Sigma = \begin{pmatrix} 1 & \dots & \dots & \dots \\ \dots & 1 & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & 1 \end{pmatrix}_{J \times J}$ (there will be $J \times (J-1)$ cov.s need to be estimated)

- Individual likelihood

$$\begin{aligned} Pr(Y_j = j | X_i = x) &= Pr(U_{ij} \geq U_{ik} \; \forall \; k | X_i = x) \\ &= Pr(\epsilon_{ij} + V_{ij} \geq \epsilon_{ik} + V_{ik} \; \forall \; k | X_i = x) \end{aligned}$$

is not available in closed form, and can be only be calculated by Monte-Carlo simulation, which is computationally infeasible due to the high dimensions of $\Sigma$

---

# Week 9: More on Multinomial Choice (Nested Logit & Random Coefficients Logit)

## *9 More On Multinomial Choice - A*

### *Nested Logit (Tractable way to include correlations in $\epsilon$ I)*

- Nested Logit

#### *Setup*

- Allows $\epsilon_{ij}$ to be correlated within nests but not across nests
- (*drop index $i$ for simplicity*)
- Still assume $\epsilon_j$ to follow Type-I Extreme Value distribution with CDF: $F(\epsilon) = \exp(-\exp(-\epsilon))$, but they can be *correlated within nests*

#### *Joint CDF*

- Joint CDF for Nested Logit with $J$ nests and $K_j$ options within each nest:

$$\begin{aligned} F(\epsilon_{11}, \dots, \epsilon_{JK_J}) &= \prod_{j=1}^{J} \exp\left(-\left[\sum_{k=1}^{K_j} \exp\left(-\frac{\epsilon_{jk}}{\tau_j}\right)\right]^{\tau_j}\right) \\ &= \exp\left(-\sum_{j=1}^{J}\left[\sum_{k=1}^{K_j} \exp\left(-\frac{\epsilon_{jk}}{\tau_j}\right)\right]^{\tau_j}\right) \end{aligned}$$

- $\tau_j$ allows $\epsilon_{jk}$ to be correlated within nest $j$ with $\rho_j = 1 - \tau_j^2$
  - $\tau_j \uparrow$ correlation $\downarrow$

    - $\tau_j = 1 \implies \epsilon_{jk}$ are independent (like in the multinomial logit) / or there is just 1 option in nest $j$
    - $\tau_j = 0 \implies \epsilon_{jk}$ are perfectly correlated
- We assume/pre-specify the nests (correlation structure) but need to estimate $\tau$'s

### *Likelihood and Estimation*

- Individual Likelihood: For $U_{jk} = V_{jk} + \epsilon_{jk}$, we have:

$$P_{jk} = P_{k|j} \times P_j$$

  where

$$\begin{cases} P_{k|j} = \dfrac{\exp\left(\frac{V_{jk}}{\tau_j}\right)}{\sum_{m=1}^{K_j} \exp\left(\frac{V_{jm}}{\tau_j}\right)} \\ P_j = \dfrac{\left(\sum_{m=1}^{K_j} \exp\left(\frac{V_{jm}}{\tau_j}\right)\right)^{\tau_j}}{\sum_{l=1}^{J} \left(\sum_{m=1}^{K_l} \exp\left(\frac{V_{lm}}{\tau_l}\right)\right)^{\tau_l}} \end{cases}$$

- $\theta$ (including $\tau_1, \ldots, \tau_J$) is estimated by MLE
- *IIA no longer holds*
- Analytically and computationally tractable
- This is not a sequential choice!

## *Random Coefficients Logit Model (Tractable way to include correlations in $\epsilon$ II)*

- Random Coefficients Logit Model

### *Setup*

- Allows for preferences $\gamma_i$ to be heterogeneous across consumers:

$$U_{ij} = W_i' \beta_j + W_{ij}' \gamma_i + \epsilon_{ij}$$

  and we need to parameterise the distribution of $\gamma_i$
    - a common choice is $\gamma_i \sim N(\gamma, \Sigma)$ (so $\lambda$ includes $\gamma, \Sigma$; $\theta$ includes $\lambda, \beta$) with $\Sigma$ restricted to be diagonal

### *Likelihood and Estimation*

- Conditional on $\gamma_i$, this is just the standard multinomial logit:

$$Pr(Y_i = j | x_i, \gamma_i; \beta) = \frac{\exp(W_i' \beta_j + W_{ij}' \gamma_i)}{\sum_{k=1}^{J} \exp(W_i' \beta_k + W_{ik}' \gamma_i)}$$

- To get $Pr(Y_i = j|x_i; \beta, \lambda)$, we integrate over $\gamma_i$:

$$Pr(Y_i = j|x_i; \beta, \lambda) = \int \underbrace{\frac{\exp(W_i'\beta_j + W_{ij}'\gamma_i)}{\sum_{k=1}^{J} \exp(W_i'\beta_k + W_{ik}'\gamma_i)}}_{Pr(Y_i=j|x_i,\gamma_i;\beta)} f(\gamma; \lambda)\, d\gamma$$

and is usually calculated by Monte-Carlo integration:

$$\frac{1}{S} \sum_{s=1}^{S} Pr(Y_i = j|X_i, \gamma_s; \beta)$$

for some large $S$, where $\gamma_s$ are iid draws from $F(\lambda)$
- $\beta, \lambda$ can be estimated by MLE with likelihood function:

$$L_n(\beta, \lambda) = \prod_{i=1}^{n} Pr(Y_i = y_i|x_i; \beta, \lambda)$$

- *IIA no longer holds*

## *Estimation with Aggregate Data*

- Estimation with Aggregate Data

### *Setup*

- $$U_{ij} = \underbrace{X_j'\gamma + \xi_j}_{\delta_j} + \epsilon_{ij}$$
    - $X_j$ includes the price and observed characteristics of product $j$
    - $\xi_j$ captures unobserved attributes of the product, unquantifiable factors, systematic shocks, etc.
    - $\epsilon_{ij}$ are iid Type I Extreme Values
- Apart from $J$ goods, there is an outside good/option with $\delta_0 = 0$

### *Estimation*

- The market shares are given by:

$$s_j = Pr(Y_i = j) = \frac{\exp(\delta_j)}{1 + \sum_{k=1}^{J} \exp(\delta_k)}$$

- Then:

$$\log(s_j) - \log(s_l) = \log\left(\frac{\frac{\exp(\delta_j)}{1+\sum_{k=1}^{J}\exp(\delta_k)}}{\frac{\exp(\delta_l)}{1+\sum_{k=1}^{J}\exp(\delta_k)}}\right) = \log\left(\frac{\exp(\delta_j)}{\exp(\delta_l)}\right) = \delta_j - \delta_l$$

1. Since (assume with no loss in generality) $\delta_0 = 0$, we can *back out $\delta_j$ from the market shares*:

$$\delta_j = \log(s_j) - \log(s_0)$$

and focus on:

$$\delta_j = X_j'\gamma + \xi_j$$

2. To *tackle with endogeneity of price* ($p_j$ is likely to be correlated with $\xi_j$), we need Instrument Variables

- This can be generalised to a Random Coefficient version, but the steps are always: 1. back out $\delta_j$ from market shares; 2. Use IV

---

## Others

- ECON0021 Week 7 Question
- Estimand is a population concept; Estimator is its finite sample analogue
- Inverse of a $2 \times 2$ matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

- Statistical independence:

$$\begin{aligned} A \perp B &\Leftrightarrow F_{A|B}(a|b) = F_A(a) \\ &\Leftrightarrow Pr(A < a|b) = Pr(A < a) \\ &\Leftrightarrow P(A = a, B = b) = P(A = a)P(B = b) \\ &\Leftrightarrow F_{AB}(a, b) = F_a(a)F_B(b) \\ &\Leftrightarrow f_{AB}(a, b) = f_A(a)f_B(b) \end{aligned}$$

- Conditional Expectation:

$$E[A|B] = \int_{-\infty}^{\infty} a f_{A|B}(a|b)\, da$$

and $f_{A|B}(a|b) = \frac{f_{A,B}(a,b)}{f_B(b)}$

- Central Limit Theorem: Let $X_i, i = 1, \ldots, n$ are random samples independently drawn from a distribution with mean $\mu$ and variance $\sigma^2$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample average, then as $n \to \infty$:

$$\sqrt{n}(\bar{X}_n - \mu) \sim^d \mathcal{N}(0, \sigma^2)$$

- Confidence Interval (95%):

$$\left[-1.96\frac{\sigma}{\sqrt{n}} + \bar{X}_n < \mu < 1.96\frac{\sigma}{\sqrt{n}} + \bar{X}_n\right]$$

- OLS Estimand in Matrices:

$$\beta = E[X'X]^{-1}E[X'Y]$$

- OLS Estimator in Matrices:

$$\hat{\beta} = \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} \left(\sum_{i=1}^{n} X_i Y_i\right)$$

- pdf of $\mathcal{N}(0, \sigma^2)$:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{x^2}{2\sigma^2})$$

- Chain Rule for Vectors

$$\frac{\partial}{\partial \beta}F(x^\intercal \beta) = xF'(x^\intercal \beta); \frac{\partial}{\partial \beta^T}F(x^\intercal \beta) = x^\intercal F'(x^\intercal \beta)$$

- Jensen Inequality
  - For $f(x) = x^2$, we have a nice proof:
    $Var(X) = E[X^2] - (E[X])^2 > 0 \implies E[X^2] > (E[X])^2$

---

## Revisions

- pdf of continuous distributed variable:

$$f_x(x) = Pr(x \in (x - \epsilon, x + \epsilon)))$$

- conditional to unconditional pdf

$$Pr(X|Y) = \frac{Pr(X,Y)}{Pr(Y)} \rightsquigarrow \text{(apply to cdf and pdf)}$$

- expectation as integrals

$$\mathbb{E}[X] = \begin{cases} \sum_k k \cdot Pr(X = k) & \text{for discrete X} \\ \rightarrow \int k \cdot f_X(x = k)\, dx & \text{as } k \rightarrow \infty \text{ for continuous X} \end{cases}$$

- conditional expectations

$$\mathbb{E}[X|Y] = \int k f_X(x = k)\, dk$$

- LIE in integrals

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]] = \int\int x f_{X|Y}(k|y)\, dx f_Y\, dy$$

- *make brief explanations of the steps*

- *make a cheatsheet of the full MLE workflow of possible distributions*
- MLE: sample likelihood -> take log -> FOC with respect to parameters -> score -> Hessian -> asymp Var
- Asymptotic var:

$$AsympVar(X) = \sqrt{n} \cdot Var(x)$$