ECON 626: Applied Microeconomics

Lecture 11:

Maximum Likelihood Estimation

Professors: Pamela Jakiela and Owen Ozier

Maximum Likelihood: Motivation

So far, we've been thinking about average treatment effects, but the ATE may or may not be the main quantity of interest research-wise

- Imperfect compliance $\Rightarrow$ LATE/TOT estimates
- Outcomes may be censored (as in a tobit model)
    - OLS estimates of the treatment effect are inconsistent
- Treatments may impact specific parameters in a structural or theoretical model; may want to know how much parameters change
    - Theory can provide a framework for estimating treatment effects

ML approaches can help to translate treatment effects into "economics"

Maximum Likelihood: Overview

In ML estimation, the data-generating process is the theoretical model

- First key decision: what is your theoretical model?
  - Examples: utility function, production function, hazard model
- Second key decision: continuous vs. discrete outcome variable
  - Censoring, extensions lead to intermediate cases
- Third key decision: structure of the error term
  - Typically additive, but distribution matters

OLS in a Maximum Likelihood Framework

Consider a linear model:

$$y_i = X_i'\beta + \varepsilon_i \text{ where } \varepsilon_i | X_i \sim \mathcal{N}(0, \sigma^2)$$

$$\Rightarrow y_i \sim \mathcal{N}(X_i'\beta, \sigma^2)$$

The normal error term characterizes the distribution of y:

$$f(y|X;\theta) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\left[\left(\frac{y - X'\beta}{\sigma}\right)^2 / 2\right]}$$

$$= \frac{1}{\sigma}\phi\left(\frac{y - X'\beta}{\sigma}\right)$$

$$= \mathcal{L}(\theta)$$

where $\theta = (\beta, \sigma)$

OLS in a Maximum Likelihood Framework

Knowing $f(y|X;\theta)$, we can write down the log-likelihood function for θ:

$$\ell(\theta) = \sum_i \ln\left[f(y_i|X_i;\theta)\right]$$

$$= \sum_i \ln\left[\frac{1}{\sigma}\phi\left(\frac{y_i - X_i'\beta}{\sigma}\right)\right]$$

ML Estimation in Stata

Estimating $\hat{\beta}$ in Stata:

```
capture program drop myols
program myols
args lnf beta sigma
quietly replace `lnf'=log((1/`sigma')*normalden(($ML_y1-`beta')/`sigma'))
end

ml model lf myols (beta: y = x) /sigma
ml maximize
```

where `$ML_y1` is the dependent variable

- By default, Stata imposes a linear structure on independent variable

Tobit Estimation

Suppose we only observe y_i^* if $y_i^* > 0$

$$C_i = \begin{cases} 0 & \text{if } y_i^* > 0 \\ 1 & \text{if } y_i^* \leq 0 \end{cases}$$

So, we observe: $\left(X_i, y_i^* \cdot (1 - C_i), C_i\right)$ for each observations i

Tobit Estimation

Suppose we only observe y_i^* if $y_i^* > 0$

$$C_i = \begin{cases} 0 & \text{if } y_i^* > 0 \\ 1 & \text{if } y_i^* \leq 0 \end{cases}$$

So, we observe: $(X_i, y_i^* \cdot (1 - C_i), C_i)$ for each observations i

With censoring of y_i^* at 0, the likelihood function takes the form:

$$\mathcal{L}_i(\theta) = [f(y_i^* | X_i; \theta)]^{1-C_i} \cdot [\Pr(y_i^* \leq 0 | X_i; \theta)]^{C_i}$$

Tobit Estimation

Since $\varepsilon_i = y_i^* - X_i'\beta$, we know that:

$$\begin{aligned}\Pr\left(y_i^* \leq 0 | X_i; \theta\right) &= \Pr\left(\varepsilon_i < -X_i'\beta\right) \\ &= \Phi\left(-\frac{X_i'\beta}{\sigma}\right) \\ &= 1 - \Phi\left(\frac{X_i'\beta}{\sigma}\right)\end{aligned}$$

Tobit Estimation

Since $\varepsilon_i = y_i^* - X_i'\beta$, we know that:

$$\Pr\left(y_i^* \leq 0 | X_i; \theta\right) = \Pr\left(\varepsilon_i < -X_i'\beta\right)$$

$$= \Phi\left(-\frac{X_i'\beta}{\sigma}\right)$$

$$= 1 - \Phi\left(\frac{X_i'\beta}{\sigma}\right)$$

We can re-write the likelihood as:

$$\mathcal{L}_i(\theta) = \left[\frac{1}{\sigma}\phi\left(\frac{y - X'\beta}{\sigma}\right)\right]^{1-C_i} \cdot \left[1 - \Phi\left(\frac{X_i'\beta}{\sigma}\right)\right]^{C_i}$$

Tobit ML Estimation in Stata

Modifying the Stata likelihood function to adjust for censoring:

```
capture program drop mytobit
program mytobit
args lnf beta sigma
quietly replace `lnf'=log((1/`sigma')*normalden(($ML_y1-`beta')/`sigma'))
quietly replace `lnf'= log(1-normal(`beta'/`sigma')) if $ML_y1==0
end

ml model lf myols (beta: ystar = x) /sigma
ml maximize
```

Why Use Maximum Likelihood?

Many economic applications start from a non-linear model of an individual decision rule some other underlying structural process

- Impacts on preferences (e.g. risk, time)
- Duration of unemployment spells
- CES production, utility functions

Maximum likelihood in Stata vs. Matlab:

- Stata is fast and (relatively) easy, if it converges
- No restrictions on the functional form of the likelihood in Matlab
- Broader range of optimization options in Matlab

Maximum Likelihood Estimation

Let y_i be the observed decision in choice situation i for $i = 1, \ldots, I$

$$y_i = g\left(x_i; \theta\right) + \varepsilon_i$$

where x_i denotes the exogenous parameters of the situation (e.g. price), θ denotes the preference/structural parameters, and $\varepsilon_i \sim \mathcal{N}(0, \sigma^s)$

- Space of outcomes/choices is continuous (i.e. not discrete)
- $g\left(x; \theta\right) + \varepsilon_j$ is the structural model (e.g. demand function)
  - Often derived by solving for utility-maximizing choice

Maximum Likelihood Estimation

Let y_i be the observed decision in choice situation i for $i = 1, \ldots, I$

$$y_i = g\left(x_i; \theta\right) + \varepsilon_i$$

where x_i denotes the exogenous parameters of the situation (e.g. price), θ denotes the preference/structural parameters, and $\varepsilon_i \sim \mathcal{N}(0, \sigma^s)$

- Space of outcomes/choices is continuous (i.e. not discrete)
- $g\left(x; \theta\right) + \varepsilon_j$ is the structural model (e.g. demand function)
  - Often derived by solving for utility-maximizing choice

Because $\varepsilon_i \sim \mathcal{N}(0, \sigma^s)$, we know that $\underbrace{y_i - g\left(x_i; \theta\right)}_{\varepsilon_i} \sim \mathcal{N}(0, \sigma^2)$

$$\Rightarrow y_i \sim \mathcal{N}(g\left(x_i; \theta\right), \sigma^2)$$

Maximum Likelihood Estimation: CRRA Example

Assume utility over income takes the constant relative risk aversion (CRRA) form given risk aversion parameter $\rho > 0$:

$$u(x) = \frac{x^{1-\rho}}{1-\rho}$$

Maximum Likelihood Estimation: CRRA Example

Assume utility over income takes the constant relative risk aversion (CRRA) form given risk aversion parameter $\rho > 0$:

$$u(x) = \frac{x^{1-\rho}}{1-\rho}$$

Agent chooses an amount, $z \in [0, b]$, to invest in a risky security that yields payoff of 0 with probability $\frac{1}{2}$ and payoff of λz with probability $\frac{1}{2}$

$$\max_{z \in [0,b]} \frac{1}{2(1-\rho)} \left[(b-z)^{1-\rho} + (b+\lambda z)^{1-\rho} \right]$$

The optimal interior allocation to the risky security is given by

$$z^* (b, \lambda) = b \left(\frac{\lambda^{1/\rho} - 1}{\lambda^{1/\rho} + \lambda} \right)$$

Maximum Likelihood Estimation: CRRA Example

People implement their choices with error:

$$z_i = z^* \left(b, \lambda\right) + \varepsilon_i \text{ where } \varepsilon_i | b, \lambda \sim \mathcal{N}(0, \sigma^s)$$

The normal error term characterizes the distribution of y:

$$f\left(z_i | b, \lambda; \theta\right) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\left[\left(\frac{y - z_i^*(b,\lambda)}{\sigma}\right)^2 / 2\right]}$$

$$= \frac{1}{\sigma}\phi\left(\frac{y - z_i^*\left(b, \lambda\right)}{\sigma}\right)$$

where $\theta = (\rho, \sigma)$

Maximum Likelihood Estimation: CRRA Example

We only observe z_i^* if $z_i^* < b$

$$C_i = \begin{cases} 0 & \text{if } z_i^* < b \\ 1 & \text{if } z_i^* \geq b \end{cases}$$

Maximum Likelihood Estimation: CRRA Example

We only observe z_i^* if $z_i^* < b$

$$C_i = \begin{cases} 0 & \text{if } z_i^* < b \\ 1 & \text{if } z_i^* \geq b \end{cases}$$

With censoring, the likelihood function takes the form:

$$\mathcal{L}_i(\theta) = \left[f\left(z_i | b, \lambda; \theta\right)\right]^{1-C_i} \cdot \left[\Pr\left(z_i \geq b; \theta\right)\right]^{C_i}$$

Maximum Likelihood Estimation: CRRA Example

We only observe z_i^* if $z_i^* < b$

$$C_i = \begin{cases} 0 & \text{if } z_i^* < b \\ 1 & \text{if } z_i^* \geq b \end{cases}$$

With censoring, the likelihood function takes the form:

$$\mathcal{L}_i(\theta) = \left[f\left(z_i|b, \lambda; \theta\right)\right]^{1-C_i} \cdot \left[\Pr\left(z_i \geq b; \theta\right)\right]^{C_i}$$

Log likelihood takes the form:

$$\ell_i(\theta) = \left(1 - C_i\right) \ln\left[f\left(z_i|b, \lambda; \theta\right)\right] + C_i \ln\left[\Pr\left(z_i \geq b|\theta\right)\right]$$

Maximum Likelihood Estimation: CRRA Example

Because we know that $\varepsilon_i | b, \lambda \sim \mathcal{N}(0, \sigma^s)$, we can calculate:

$$\begin{aligned}
\Pr\left(z_i \geq b | \theta\right) &= \Pr\left(z_i^*(b, \lambda) + \varepsilon_i \geq b | \theta\right) \\
&= 1 - \Pr\left(\varepsilon_i \leq b - z_i^*(b, \lambda) | \theta\right) \\
&= 1 - \Phi\left(\frac{b - z_i^*(b, \lambda)}{\sigma}\right)
\end{aligned}$$

Maximum Likelihood Estimation: CRRA Example

Because we know that $\varepsilon_i | b, \lambda \sim \mathcal{N}(0, \sigma^s)$, we can calculate:

$$\begin{aligned}
\Pr\left(z_i \geq b|\theta\right) &= \Pr\left(z_i^*(b,\lambda) + \varepsilon_i \geq b|\theta\right) \\
&= 1 - \Pr\left(\varepsilon_i \leq b - z_i^*(b,\lambda)|\theta\right) \\
&= 1 - \Phi\left(\frac{b - z_i^*(b,\lambda)}{\sigma}\right)
\end{aligned}$$

We can re-write the log likelihood as:

$$\ell_i(\theta) = (1 - C_i)\ln\left[\frac{1}{\sigma}\phi\left(\frac{y - z_i^*(b,\lambda)}{\sigma}\right)\right] + C_i \ln\left[1 - \Phi\left(\frac{b - z_i^*(b,\lambda)}{\sigma}\right)\right]$$

Maximum Likelihood Estimation: CRRA Example

Stata program for ML estimation in a non-linear framework:

```
capture program drop mymodel
program mymodel

args lnf rho sigma
tempvar ratio res

quietly gen double ‘ratio’ = $ML_y2*(($ML_y3^(1/‘rho’) - 1)/($ML_y3^(1/‘rho’) +
quietly gen double ‘res’ = $ML_y1 - ‘ratio’

quietly replace ‘lnf’ = ln((1/‘sigma’)*normalden((‘res’)/‘sigma’))
quietly replace ‘lnf’= ln(1-normal(($ML_y2-‘ratio’)/‘sigma’)) if $ML_y4==1

end

ml model lf mymodel (rho: investment budget return censor = ) (sigma: )
ml maximize
```