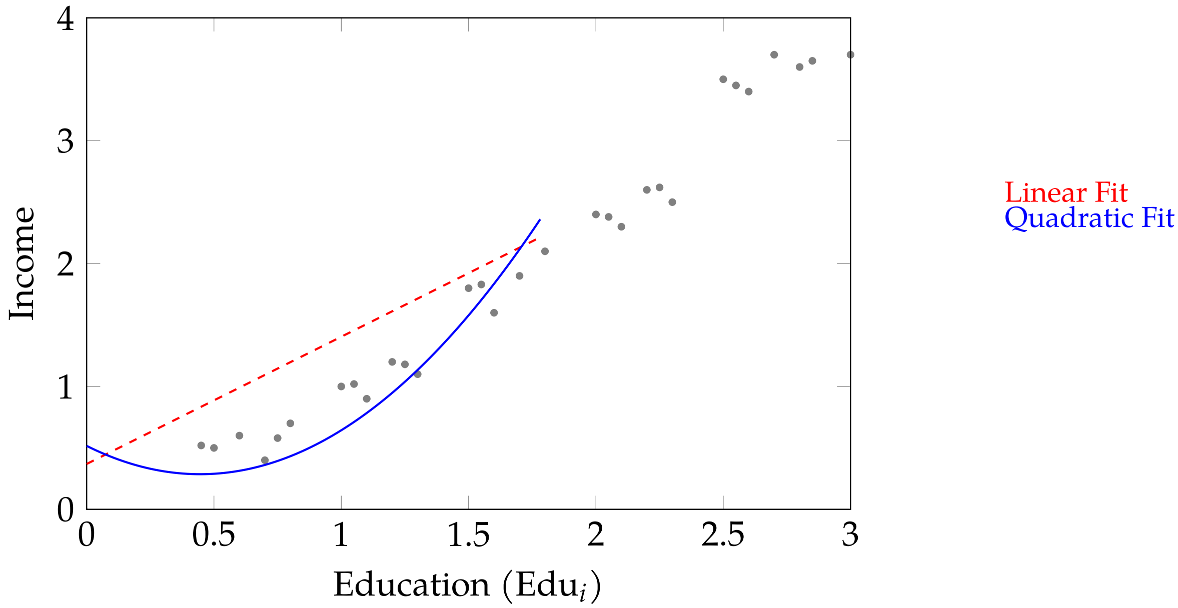
ECON 640 Stationary Non-Linear Models

Dr. Yang Liang

1 Quadratic Regression Model (Semi-parametric)

$$\text{Income} = \beta_1 \text{Edu}_i + \beta_2 \text{Edu}_i^2 + u_i$$



General Model Formulations

$$y_i = \sum X_i(\beta) + u_i$$
$$y_i = f(X_i) + u_i$$

Polynomials

General form:

$$y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_k X_i^k + u_i$$

2 The Exponential Function & Natural Logarithms

$$\ln\left(\frac{1}{x}\right) = -\ln(x)$$
$$\ln(ax) = \ln(a) + \ln(x)$$
$$\ln\left(\frac{x}{a}\right) = \ln(x) - \ln(a)$$
$$\ln(x^d) = d\ln(x)$$

Logarithms and Percentages

When Δx is small:

$$\ln(x + \Delta x) - \ln(x) \approx \frac{\Delta x}{x}$$

If $x = 100$ and $\Delta x = 1$, this represents a log points difference.

Three Logarithmic Regression Models

Case I: Linear-Log Model

$$y_i = \beta_0 + \beta_1 \ln(x_i) + u_i$$

Interpretation:

- A 1% change in x is associated with a change in y of $\frac{\beta_1}{100}$.

Case II: Log-Linear Model

$$\ln(y_i) = \beta_0 + \beta_1 x_i + u_i$$

Interpretation:

- A one-unit change in x ($\Delta x = 1$) is associated with a $(100 \times \beta_1)\%$ change in y.

Approximation for percentage change:

$$\Delta y / y \approx \beta_1 \Delta x$$

Or:

$$100 \times (e^{\beta_1} - 1)$$

Case III: Log-Log Model

$$\ln(y_i) = \beta_0 + \beta_1 \ln(x_i) + u_i$$

Interpretation:

- β_1 represents elasticity: the percentage change in y per percentage change in x.

Example: Log-Linear Model Example

$$\ln(y_i) = \beta_0 + \beta_1 x_i + u_i$$

Example:

- x_i: Policy (e.g., Recreational Marijuana Laws)
- y_i: Tobacco use in the last 30 days
- $\hat{\beta}_1 = -0.110$

Interpretation of β_1:

$$\ln(y_2) - \ln(y_1) = \hat{\beta}_1$$
$$\frac{y_2}{y_1} = e^{\hat{\beta}_1}$$
$$\Delta y / y = e^{\hat{\beta}_1} - 1$$

For $\hat{\beta}_1 = -0.110$:

$$1 - e^{-0.110} \approx 0.1047 \quad (10.47\% \text{ decrease in } y)$$

If positive:

$$e^{0.110} - 1 \approx 11.6\%$$

3 Binary Dependent Variable

Question: Study whether race is a determining factor in a mortgage application.

$$\text{mortgage success}_i = \alpha + \beta \frac{P}{I}_i + u_i$$

$$\text{where, payment-to-income ratio: } \frac{P}{I}$$

you can estimate this using:

- Linear Probability Model (LPM)
- Probit or Logit

Introduction to MLE

Maximum Likelihood Estimation (MLE):

- **Objective:** MLE is a method to estimate parameters of a statistical model by finding the values that maximize the likelihood of the observed data.
- *Intuition:* Consider guessing the mean height of a group of people. If $Y_i \sim N(\mu, \sigma^2)$, then:
  - Observe a random sample.
  - Estimate the sample mean as a way to maximize the probability of the observed data.

Example 1: Estimating the Mean from a Normal Distribution

Assume $Y_i \sim N(\mu, \sigma^2)$.

Probability Density Function (PDF):

$$f(y_i|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}$$

Likelihood Function:

$$L(\mu) = \prod_{i=1}^{n} f(y_i|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}$$

Log-Likelihood Function:

$$\ln L(\mu) = \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2$$

First-Order Condition (FOC):

$$\frac{\partial \ln L(\mu)}{\partial \mu} = -\frac{1}{\sigma^2} \sum (y_i - \mu) = 0$$

Result:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} y_i \quad \text{(sample mean)}$$

Example 2: Using MLE to Derive OLS Estimators

Consider the linear regression model:

$$y_i = \alpha + \beta x_i + u_i, \quad u_i \sim N(0, \sigma^2)$$

Assumption: To use MLE, we need to assume some functional form for the error term – we pick normality here.

Likelihood Function:

$$L(\alpha, \beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \alpha - \beta x_i)^2}{2\sigma^2}}$$

Log-Likelihood Function:

$$\ln L(\alpha, \beta) = \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

First-Order Conditions:

$$\frac{\partial \ln L}{\partial \alpha} = -\frac{1}{\sigma^2} \sum (y_i - \alpha - \beta x_i) = 0$$

$$\frac{\partial \ln L}{\partial \beta} = -\frac{1}{\sigma^2} \sum (y_i - \alpha - \beta x_i)(-x_i) = 0$$

Solution:

$$\hat{\beta} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

Example 3: Logit Model

For binary dependent variables, consider the logit model:

$$P(Y_i = 1 | X_i) = \frac{1}{1 + e^{-\alpha - \beta X_i}}$$

Probability Functions:

$$P(Y_i = 0 | X_i) = 1 - P(Y_i = 1 | X_i)$$

Likelihood Function:

$$L(\alpha, \beta) = \prod_{i=1}^{n} P(Y_i = 1 | X_i)^{Y_i} (1 - P(Y_i = 1 | X_i))^{1 - Y_i}$$

Log-Likelihood Function:

$$\ln L(\alpha, \beta) = \sum_{i=1}^{n} Y_i \ln(P_i) + (1 - Y_i) \ln(1 - P_i)$$

FOC:

$$\frac{\partial \ln L}{\partial \alpha} = 0, \quad \frac{\partial \ln L}{\partial \beta} = 0$$

Note: Analytical solutions are generally not available for logit models, so numerical methods are used for estimation.

What Do We Do After Numerically Solving β ? $\rightarrow$ Marginal Effects

Once we obtain the estimates of β (e.g., from a logit model), we often calculate **marginal effects** to understand the impact of a unit change in an explanatory variable on the predicted probability.

4 Marginal Effects for Logit Model

Probability in Logit Models

The probability of $Y = 1$ (success) is given by:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\alpha+\beta X)}} = P$$

Deriving the Marginal Effect

The marginal effect measures the change in probability P with respect to a change in the predictor X:

$$\frac{\partial P}{\partial X} = \frac{\partial}{\partial X}\left(\frac{1}{1 + e^{-(\alpha+\beta X)}}\right)$$

Using the chain rule:

- Let $g(X) = 1 + e^{-(\alpha+\beta X)}$,
- Then:
$$\frac{\partial P}{\partial X} = \frac{1}{[g(X)]^2} \cdot \frac{\partial g(X)}{\partial X}$$
- The derivative of $g(X)$ is:
$$\frac{\partial g(X)}{\partial X} = -\beta e^{-(\alpha+\beta X)}$$

Simplifying:

$$\frac{\partial P}{\partial X} = P(1 - P) \cdot \beta$$

Key Points to Note

- **Nonlinear Effects:** The marginal effect depends on both P and β, meaning it varies across values of X.
- **Maximum Marginal Effect:** The marginal effect is largest when $P = 0.5$, as $P(1 - P)$ is maximized at this point.
- **Interpretation:**
  - At specific values of X, compute P and use $P(1 - P)\beta$ to find the marginal effect.
  - Marginal effects provide a clear interpretation of how changes in X impact the probability of success.

Practical Steps to Compute Marginal Effects

1. Estimate the coefficients $\hat{\alpha}$ and $\hat{\beta}$ from the logit model. 2. Compute the predicted probability P for each observation:

$$P = \frac{1}{1 + e^{-(\hat{\alpha}+\hat{\beta}X)}}$$

3. Use the marginal effect formula:

$$\frac{\partial P}{\partial X} = \hat{\beta} \cdot P(1 - P)$$

4. Average the marginal effects across observations or compute them at specific values of X.

Conclusion: Why Marginal Effects Matter

Marginal effects are crucial for interpreting the practical implications of models like logit and multinomial logit:

- For policy analysis, marginal effects show how changes in predictors (e.g., income, education) influence outcomes (e.g., loan approval, transportation choice).
- They allow us to compare the relative importance of predictors in models with categorical outcomes.

5 Multinomial Logit Model (MNL)

When to Use?

The multinomial logit model is used when:

- The dependent variable is **categorical** and has **more than two unordered categories**.

  Examples of Choice Sets:

- Transportation mode: {car, bus, bike}
- Housing type: {apartment, house, condo}
- Political candidate: {Trump, Harris, etc.}

Simple Model Specification

Let Y have J categories. We must:

- Set one category as the **base/reference category**.
- Estimate the probabilities for each of the other $J - 1$ categories relative to the base category.

  The probability of choosing category j (for $j = 1, 2, \dots, J$) is:

$$P(Y = j|X) = \frac{e^{\alpha_j+\beta_j X}}{1 + \sum_{k=1}^{J-1} e^{\alpha_k+\beta_k X}}$$

where:

- α_j is the intercept for category j,
- β_j represents the coefficients for the predictors X,
- The denominator ensures that probabilities sum to 1 across all J categories.

For the Reference Group (Base Category):

$$P(Y = \text{base}|X) = \frac{1}{1 + \sum_{k=1}^{J-1} e^{\alpha_k + \beta_k X}}$$

Interpreting Marginal Effects

Marginal effects for the multinomial logit model show how a one-unit change in a predictor X affects the probability of selecting a particular category j. These effects are calculated for each non-base category relative to the base category. They are often computed at the mean or other relevant values of X to provide actionable insights.