# ECN 565: Data Science & Econometrics II

Alvin Murphy

Lecture 8a: Limited Dependent Variables – Binary Choice

# Why limited dependent variables?

- To date, explicitly/implicitly assuming dependent variable was:
  - Continuous
  - Infinite support
  - Minor exception: Wages, for example, not negative
- Many economic outcomes are not continuous real numbers:
  - binary choices, ordered categories, counts, corner solutions, discrete choices.

# Examples of limited dependent variables

- Binary: Attend University , enter labor force
- Ordered: credit ratings, product star ratings.
- Count: number of doctor visits, number of arrests
- Corner solutions: cigarettes per day, hours worked per week (with many zeros).
- Multinomial discrete choice: transport mode, car choice, neighborhood choice

# Goals:

- Goal: understand the causes/determinants of economic outcomes
  - That's always our goal!
- CLM may be incapable of providing a sensible framework
  - CLM: Classical Linear Model
- Use models that respect economic constraints on outcome variable
  - Give interpretable parameters and probabilities.
- Limited dependent variable models use a nonlinear approach to estimation
  - MLE: Maximum Likelihood Estimation
  - More complicated than OLS.

# Linear Probability Model (LPM)

- Begin by considering alternative to non-linear model
- Binary outcome, so dependent variable always equals 0 or 1
- Linear Probability Model/LPM:

$$y_i = X_i'\beta + u_i, \text{ where } y_i \in \{0, 1\}.$$

- OLS estimate $\hat{\beta} = [\sum_i x_i x_i']^{-1} \sum_i x_i y_i$
- Predicted value $\hat{y}_i = X_i'\hat{\beta}$ interpreted as estimated probability.

# Limitations of LPM

- Recall mechanics of OLS
  - Choose $\hat{\beta}$ to minimize SSR
  - Residual still given by $\hat{u}_i = y_i - \hat{y}_i$
- Predicted probabilities can lie outside $[0, 1]$.
- Constant marginal effects across $X$ often unrealistic.
- Constant marginal effects across $X$ always easily interpretable
- Homer Simpson:

  "To linearity! The cause of, and solution to, all of life's problems"

# Simple Solution

- Modify model:

$$P(y_i = 1) = F(X_i'\beta), \text{ where } y_i \in \{0, 1\}.$$

- Choose $F$ function so that $0 \leq F(z) \leq 1 \; \forall z$
  - $F$ called a link function
  - Any CDF would satisfy condition that $0 \leq F(z) \leq 1 \; \forall z$
  - If $F(z)$ is normal $\rightarrow$ Probit Model
  - If $F(z)$ is logistic $\rightarrow$ Logit Model
- Certainly mechanically solves problem.

# Simple Solution

- Use of link function mechanically solves problem.
- But can we:
  - Justify economically?
  - Understand/interpret results for marginal effects?
  - Justify statistically?
  - Estimate?
- Recall our goal is to learn how $X$ affects $y$
  - we modify slightly to: learn how $X$ affects $Prob(y = 1)$

# Economic Justification: latent variable interpretation

- Introduce latent variable $y_i^*$:

$$y_i^* = X_i'\beta + u_i.$$

- Observed outcome:
  - $y_i = 1$ if $y_i^* > 0$
  - $y_i = 0$ if $y_i^* \leq 0$
- $P(y_i = 1 \mid X_i) = P(y_i^* > 0) = P(X_i'\beta + u_i > 0)$
- $P(y_i = 1 \mid X_i) = P(-u_i < X_i'\beta) = F(X_i'\beta)$ where $F$ is CDF of $-u_i$.
  - Assume symmetry, so CDF of $-u_i$ = CDF of $u_i$
  - Implicitly assumed some technical assumptions that we'll return to

# Logit and Probit: definitions

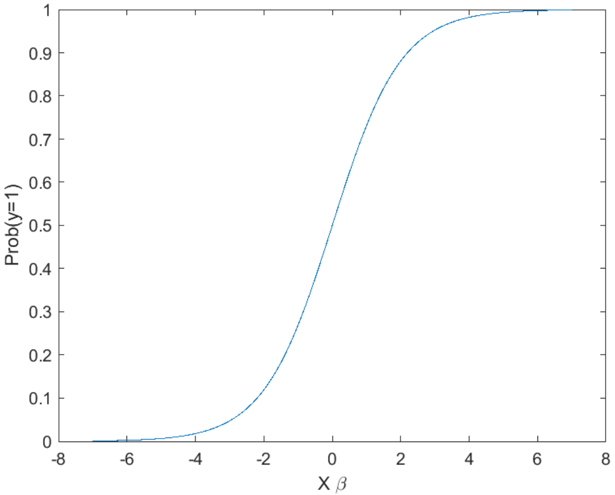
- The shape of the $F(X_i'\beta)$ function depends on the distribution of $u$
- Logit model: $u \sim logistic$: $F(X_i'\beta) = \Lambda(X_i'\beta) = \dfrac{exp(X_i'\beta)}{1+\exp(X_i'\beta)}$.
- Probit model: $u \sim normal$: $F(X_i'\beta) = \Phi(X_i'\beta) = \int_{-\infty}^{X_i'\beta} \phi(t)\,dt$,

  $\phi(t) = (2\pi)^{-1/2} e^{-t^2/2}$.
- Both options restrict $Prob(y = 1)$ to [0,1]
- Both options produce S-shaped probability curves.
- Logit has a closed-from solution. Probit does not.
  - Even with modern computers, this can matter

# Economic Justification: latent variable interpretation

- Benefit of latent variable formulation is that it can represent utility maximization
- Individual chooses $y = 1$ if utility is maximized by $y = 1$
- Example: Labor Force Participation
- Let $U_{1,i} = X_i'\beta + u_i$, $U_{0,i} = 0$ (normalize baseline).
- $y_i^* = U_{1,i} - U_{0,i} = X_i'\beta + u_i$.
- Choose $y = 1$ iff $U_1 > U_0 \Leftrightarrow X'\beta + u > 0$.
- If $u$ has logistic (normal) distribution, we obtain logit (probit) choice probability.

# Logistic CDF



# Marginal effects

- Marginal Effects are very straightforward in LPM:
  - $P(y_i = 1) = X_i'\beta$, $\frac{\partial P}{\partial x_j} = \beta_j$
  - constant marginal effect does not depend on level of $x_j$ or any $x_k$
- Richer/more complicated in logit/probit:
  - $P(y_i = 1) = F(X_i'\beta)$, $\frac{\partial P}{\partial x_j} = f(X'\beta)\beta_j$ where $f = F'$.
  - Marginal effects vary with level of $x_j$ and all $x_k$

# Evaluating and reporting marginal effects

- At what value of $X$ should we report $\frac{\partial P}{\partial x_j}$?
- Marginal effect at sample mean (MEM): evaluate $f(X'\hat{\beta})\hat{\beta}_j$ at $\bar{X}$.
- Average marginal effect (AME): $\frac{1}{n}\sum_i f(X_i'\hat{\beta})\hat{\beta}_j$.

# Evaluating and reporting effects for discrete changes

- Discrete changes often used for discrete regressors:

  $P(y = 1 \mid x_j = 1, \ldots) - P(y = 1 \mid x_j = 0, \ldots)$.

- For example: suppose $y$ and $Prob(y = 1)$ depends on two variables, $x_1$ and $x_2$
- The values of $x_1$ and $x_2$ matter. Let's set $x_2$ at its average: $x_2 = \bar{x}_2$
- Calculate the change in $P$ from increasing $x_1$ from 4 to 5
- $\Delta P = Prob(y = 1 | x_1 = 5, x_2 = \bar{x}_2) - Prob(y = 1 | x_1 = 4, x_2 = \bar{x}_2)$

# Evaluating and reporting effects for discrete changes

- $\Delta P = Prob(y = 1|x_1 = 5, x_2 = \bar{x}_2) - Prob(y = 1|x_1 = 4, x_2 = \bar{x}_2)$
- For Logit: $\Delta P = \frac{exp(\beta_0+\beta_1 5+\beta_2\bar{x}_2)}{1+exp(\beta_0+\beta_1 5+\beta_2\bar{x}_2)} - \frac{exp(\beta_0+\beta_1 4+\beta_2\bar{x}_2)}{1+exp(\beta_0+\beta_1 4+\beta_2\bar{x}_2)}$
- Non-linearity means that
  - Answer would be different if we had changed $x_1$ from 1 to 2
  - Answer would be different if we fixed $x_2$ at any other number
- A similar process applies to Probit but without a closed-from solution
- Software (e.g., Python or Stata) will calculate marginal effects
- Probit and Logit models tend to produce similar marginal effects
  - but not similar $\hat{\beta}$

# Assumptions required for estimation and inference

- Analogous to MLR1-MLR6:
    - Model is correctly specified
    - Random sample from the population
    - Conditional variation in each explanatory variable
    - $u$ and $X$ are independent
    - Homoskedasticity (or robust standard errors)
    - Assumption about the error distribution (e.g. normal, logistic)

# MLE: general idea

- Maximum Likelihood Estimation intuition:
  - choose values for $\hat{\beta}$ that maximize the likelihood of observing the outcomes in data.
- Given a density/probability $f(y \mid x, \theta)$, the likelihood is $L(\theta) = \prod_i f(y_i \mid x_i, \theta)$.
- Maximize $\ell(\theta) = \log L(\theta)$ to obtain $\hat{\theta}$.
- Under regularity conditions $\hat{\theta}$ is consistent, asymptotically normal, and efficient.

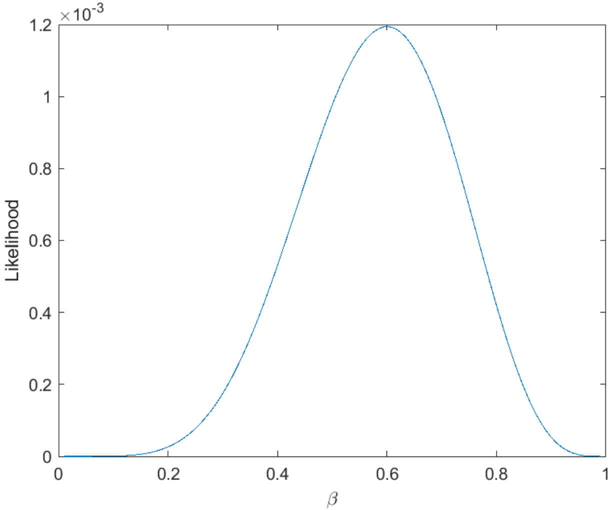
# MLE: very simple example

- We have a coin and want to know if it is fair
- Model: $Prob(y = 1) = \beta$
  - Estimate $\beta$
  - No X
- Observe data of 10 coin flips
  - $\{y_i\} = \{1, 1, 0, 1, 0, 0, 0, 1, 1, 1\}$
- $\{L_i\} = \{\beta, \beta, 1 - \beta, \beta, 1 - \beta, 1 - \beta, 1 - \beta, \beta, \beta, \beta\}$
- $L = \prod_i^n L_i = \beta^6 (1 - \beta)^4$
- $\ell(\beta) = \log L(\beta) = 6 log(\beta) + 4 log(1 - \beta)$

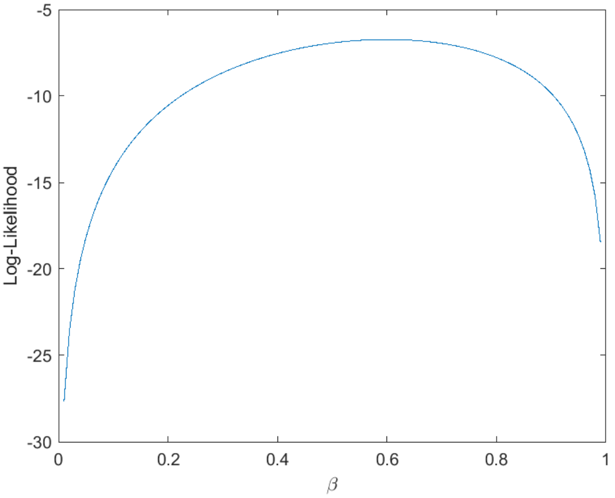
# MLE: very simple example

- $\ell(\beta) = \log L(\beta) = 6log(\beta) + 4log(1-\beta)$
- $\hat{\beta} = argmax_{\hat{\beta}}\ell(\hat{\beta}) = argmax_{\hat{\beta}}6log(\hat{\beta}) + 4log(1-\hat{\beta})$
- First order condition:
- $\frac{\partial\ell(\hat{\beta})}{\partial\hat{\beta}} = \frac{6}{\hat{\beta}} - \frac{4}{1-\hat{\beta}} = 0$
- $4\hat{\beta} = 6(1-\hat{\beta})$
- $10\hat{\beta} = 6$
- $\hat{\beta} = 0.6$

# MLE: very simple example – Likelihood function



# MLE: very simple example – Log-Likelihood function



# Logit/Probit likelihood (binary)

- Bernoulli likelihood for observation $i$:

$$P(Y = y_i) = F(X_i'\beta)^{y_i}(1 - F(X_i'\beta))^{1-y_i}$$

- if $y_i = 1$, then $L_i = F(X_i'\beta)$
- if $y_i = 0$, then $L_i = 1 - F(X_i'\beta)$
- Log-likelihood:

$$\ell(\beta) = \sum_{i=1}^{n} \left[ y_i \ln F(X_i'\beta) + (1 - y_i) \ln(1 - F(X_i'\beta)) \right].$$

# Computation and convergence practicalities

- Can't use a first order condition to solve for $\hat{\beta}$
  - Have to numerically search for $\hat{\beta}$
  - Hill climbing, quasi-newton algorithms
- Start values: zeros or LPM estimates often work.
- Numerical issues can cause divergence
  - Check for large coefficients and warnings.

# MLE – postestimation

- Hypothesis testing for MLE
  - z-tests (similar to OLS t-tests)
- Fitted values – yields probabilities that the outcome will be 1
- Goodness of Fit – pseudo R-squared
  - Pseudo $R^2 = 1 - \frac{\ell_{UR}}{\ell_0}$
  - $\ell_0$ is value for the log-likelihood with intercept only
  - As $\ell_{UR}$ increases, $\ell_{UR} \to 1$, which implies $\frac{\ell_{UR}}{\ell_0} \to 0$ and Pseudo $R^2 \to 1$
  - As $\ell_{UR}$ decreases, $\ell_{UR} \to \ell_0$, which implies $\frac{\ell_{UR}}{\ell_0} \to 1$ and Pseudo $R^2 \to 0$

# Recall our assumptions:

- Model is correctly specified
- Random sample from the population
- Conditional variation in each explanatory variable
- $u$ and $X$ are independent
- Homoskedasticity (or robust standard errors)
- Assumption about the error distribution (e.g. normal, logistic)