1 Qualitative Response Models

Qualitative response models...

- explain a "yes" or "no" choice.
- have the regressand as a dummy variable.
- are typically not estimated with OLS techniques.
- require maximum likelihood estimation.
- include linear probability (LP), logit and probit models.

1.1 Linear Probability (LP) Model

The LP model uses the regression equation

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

and ordinary least squares (OLS) to estimate the model. The conditional expectation of Y_i:

$$E(Y_i|X_i) = \beta_1 + \beta_2 X_i = P_i$$

can be treated as a probability.

Issues with the LP model:

- error terms (u_i) follow a Bernoulli rather than a normal distribution
- model exhibits heteroscedasticity.
- predicted probabilities might not satisfy $0 \leq P_i \leq 1$.
- traditional R^2 value does a poor job of measuring goodness of fit.

1.2 Logit and Probit Models

The logit and probit models specify

$$Y_i^* = \beta_1 + \beta_2 X_i + u_i$$

where Y_i^* is a latent indicator variable.

- We assume that $Y_i = 1$ when $Y_i^* > 0$.

$$\begin{aligned}
\Pr(Y_i &= 1) = \\
\Pr(Y_i^* &> 0) = \\
\Pr(\beta_1 + \beta_2 X_i + u_i &> 0) = \\
\Pr(u_i &> -\beta_1 - \beta_2 X_i) = \\
\Pr(u_i &< \beta_1 + \beta_2 X_i) = \\
\Pr(u_i/\sigma &< \beta_1/\sigma + (\beta_2/\sigma) X_i) = \\
\Pr(u_i^* &< \beta_1^* + \beta_2^* X_i) = F(\beta_1^* + \beta_2^* X_i)
\end{aligned}$$

- $\Pr(Y_i = 0) = 1 - F(\beta_1^* + \beta_2^* X_i)$
- Use maximum likelihood (ML) techniques to estimate parameters.
- When u_i is distributed normal $\rightarrow$ probit model
- When u_i is distributed logistically $\rightarrow$ logit model
- β_1^* and β_2^* are estimated, not β_1 and β_2
- Marginal effect, $\partial F(\cdot)/\partial X_i = f(\cdot)\beta_2^*$, is often reported
- Goodness of fit:
  - McFadden's pseudo R^2
  - 2×2 table of correct and incorrect predictions

1.3 Possible Applications

- Graduate student success (logit/probit model)
- Long-distance MPB dispersal (logit model)
- UW scholarship reform (probit model)
- "America the Beautiful" national park pass (interval regression model)