Probit and Logit Models

Abhilasha, Prerna, Sharada, Shreya

April 19, 2013

Motivation

- In the linear regression model, the dependent variable is quantitative. We discuss a model wherein the regressand is qualitative in nature.
- In a model where Y **is quantitative** the objective is to estimate its expected value given the value of the regressors whereas when Y **is qualitative** the objective is to find the probability of occurrence of an event. For ex: A student getting admission in IGIDR
- Qualitative response regression models are also known as **Probability Models**.

Introduction

- We study the impact of variables such as GRE, GPA and prestige of an undergraduate institute on admission into graduate school.
- Thus the response variable 'Admit' is a binary variable taking two values, 1 for admission and 0 otherwise.
- There are 3 ways to develop a probability model for a binary response variable:
  1. Linear Probability Model
  2. The Logit Model
  3. The Probit Model
- We will elucidate on the Logit and Probit models.

Probit and Logit Model

- $E(Y_i|X_i)$ is the probability that the event will occur.
- We model this probability of occurrence as the linear combination of characteristics of the individual i.e.

$$p_i = F(\beta' X_i)$$

- The probability distribution of U_i in case of a linear probability model is

U_i	P_{U_i}
$1 - \beta' X_i$	$\beta' X_i$
$-\beta' X_i$	$1 - \beta' X_i$

- Note that the distribution of U_i is not normal, it depends on the unknown parameter β and exhibits heteroskedasticity leading to the problem of interpretation.

Probit and Logit Functions

- We want a functional form such that:

$$-\infty \leq \beta' X_i \leq +\infty$$

$$0 \leq F(\beta' X_i) \leq 1$$

- Probability distribution functions satisfy the above two conditions.
- Mentioned below are the distribution functions of the Probit and Logit models

$$F(\beta' X_i) = \int_{-\infty}^{\beta' X_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}U^2} dU \equiv \Phi(\beta' X_i) \Rightarrow \text{Probit}$$

$$F(\beta' X_i) = \frac{e^{\beta' X_i}}{1 + e^{\beta' X_i}} \Rightarrow \text{Logit}$$

Latent Variable

- Consider a response function $Y_i^* = \beta' X_i + U_i$
- Latent Variable cannot be observed directly but is inferred from an observable variable.
- With reference to our example the latent variable is the utility derived from getting admission into IGIDR.

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

Odds Ratio

- In the Logit model p_i is non linear in X and β.
- p_i is the probability of getting admission (success).
- ln of Odds Ratio (logit) is linear in X and β.

$$p_i = \frac{e^{\beta' X_i}}{1 + e^{\beta' X_i}}$$

$$\text{Odds Ratio} = \frac{p_i}{1 - p_i} = e^{\beta' X_i}$$

- βk gives the change in the logit of Y (admit) for a unit change in GRE, GPA or Prestige of Institution, X_{ik} which is independent of the value of the regressor.

Marginal Effects

$$\text{Logit Model} = \frac{\partial p_i}{\partial X_{ik}} = \left[\frac{e^{\beta' X_i}}{(1+e^{\beta' X_i})^2}\right](\hat{\beta}_k)$$

$$\text{Probit Model} = \frac{\partial p_i}{\partial X_{ik}} = \phi(\hat{\beta}' X_i)(\hat{\beta}_k)$$

- Slope coefficients have a different meaning as compared to CLRM and LPM.
- The rate of change in probability with respect to X in:
  1. Logit: involves not only β but also the level of p_i from which the change is measured.
  2. Probit: depends on the level of X

Estimation

Inconsistency of OLS for Probit and Logit Models

$$p_i = F(\beta' X_i)$$

$$Y_i = E(Y_i|X_i) + \eta_i \text{ with } E(\eta_i|X_i) = 0$$

Rewrite this as,

$$Y_i = \beta' X_i + [E(Y_i|X_i) - \beta' X_i] + \eta_i$$

And let,

$$[E(Y_i|X_i) - \beta' X_i] + \eta_i = \omega_i$$

$\therefore$

$$Y_i = \beta' X_i + \omega_i$$

$$E(\omega_i|X_i) = F(\beta' X_i) - \beta' X_i$$

$\therefore$

$$E(\omega_i|X_i) = 0 \text{ only if } F(\beta' X_i) = \beta' X_i$$

Estimation

Estimation of Logit with WLS

$$L_i = \ln\left[\frac{p_i}{1-p_i}\right] = \beta' X_i + \eta_i$$

- For Individual Data OLS is infeasible as the Likelihood Function is not defined
- For Grouped Data WLS is applicable because although the error is heteroscedastic it is still normally distributed.

$$\eta_i \sim \mathcal{N}\left[0, \frac{1}{N_i p_i (1-p_i)}\right]$$

Estimation

1. Estimate for each X_i the probability of getting admission as $\hat{p}_i = \dfrac{n_i}{N_i}$.
2. Obtain the logit as $\hat{L}_i = \ln\left[\dfrac{p_i}{(1-p_i)}\right]$.
3. Transform the model to overcome heteroscedasticity as follows,
$$\sqrt{w_i}L_i = \beta_1\sqrt{w_i} + \beta_2\sqrt{w_i}X_i + \sqrt{w_i}\eta_i$$
4. Estimate this model with OLS (regression through origin).

Non Linear Least Squares

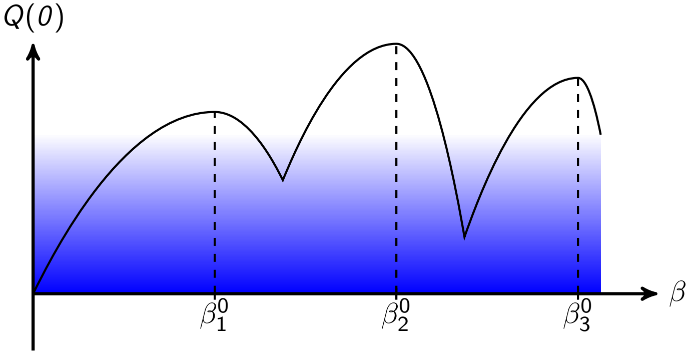
- Under the normality assumption of the disturbance term, OLS estimators are BLUE as well as BUE.
- Dropping the normality assumption on the noise term, it is possible to obtain nonlinear estimators that perform better than OLS estimators.
- Nonlinear least squares estimation involves solving nonlinear normal equations.
- Analytical solution not possible - iterative numerical search procedure needed.

$$\text{Minimise Q} = \sum \hat{\eta}_i^2 = \sum [Y_i - F(\beta' X_i)]^2$$

NLLS Contd...

- Use Newton-Raphson Method to iterate and obtain:

$$\hat{\beta}_{j+1} = \hat{\beta}_j - \left[\frac{\partial^2 Q}{\partial\beta\partial\beta'}\right]^{-1}_{\hat{\beta}_{j-1}} \left[\frac{\partial Q}{\partial\beta}\right]_{\hat{\beta}_{j-1}}$$



Maximum Likelihood Estimation

- The **maximum likelihood principle** is applicable when the form of the probability distribution is known.
- Produces estimates that are **consistent** and, at least asymptotically, minimum variance.
- Likelihood Function

$$L = \prod_{i=1}^{n} p_i^{Y_i}(1 - p_i)^{1-Y_i}$$

Likelihood Function

- Probit

$$p_i = \Phi(\beta' X_i)$$

$$L = \prod_{i=1}^{n} \Phi(\beta' X_i)^{Y_i} (1 - \Phi(\beta' X_i))^{1-Y_i}$$

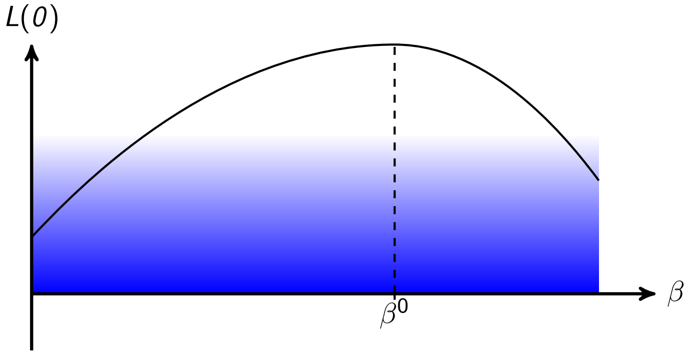
- Logit

$$p_i = \frac{e^{\beta' X_i}}{1 + e^{\beta' X_i}}$$

$$L = \prod_{i=1}^{n} \left[\frac{e^{\beta' X_i}}{1 + e^{\beta' X_i}} \right]^{Y_i} \left[\frac{1}{1 + e^{\beta' X_i}} \right]^{1-Y_i}$$

Maximum Likelihood Contd...

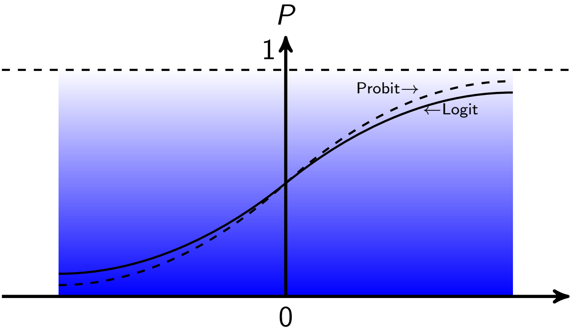
- As with nonlinear estimation closed form solution is difficult
- Iterative procedures (Newton Raphson) must be used (replace Q with $\log L$).
- Convergence to global maximum is ensured as Log function is globally concave



Choice between Probit and Logit

- Logistic Distribution has flatter tails compared to Probit
- The Logit model performes well in heterogenous data, moderately balanced data as well as data with outliers
- Logit and Probit estimates are approximately related by the following rule:

$$\widehat{\beta_{Logit}} = 1.6\widehat{\beta_{Probit}}$$



Example

- The example attempts to model various factors that influence the admission of a student into a graduate school.
- The dependent variable; Admit/Dont Admit is binary.
- The explanatory variables are:
  1. GRE score - treated as a continuous variable
  2. GPA - treated as a continuous variable
  3. Prestige of the undergraduate institution - coded as rank, taking values from 1 to 4 where Rank 1 has the highest prestige.

SAS Logit Output

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Std Est	Wald Error	χ^2	$Pr > \chi^2$
Intercept	1	-5.5414	1.1381	23.7081	< 0.0001
GRE	1	0.00226	0.00109	4.2842	0.0385
GPA	1	0.8040	0.3318	5.8714	0.0154
Rank 1	1	1.5514	0.4178	13.7870	0.0002
Rank 2	1	0.8760	0.3667	5.7056	0.0165
Rank 3	1	0.2112	0.3929	0.2891	0.5908

Logit Interpretation

- For every one unit change in GRE, the log odds of admission (versus non-admission) increases by 0.002.
- For a one unit increase in GPA, the log odds of being admitted to graduate school increases by 0.804
- The coefficients for the categories of rank have a slightly different interpretation because rank is a qualitative variable.

Odds Ratio Estimates

Odds Ratio Estimates			
Effect	Point Est	95% Wald UL	95% Wald LL
GRE	1.002	1.000	1.004
GPA	2.235	1.166	4.282
RANK 1 vs 4	4.718	2.080	10.701
RANK 2 vs 4	2.401	1.170	4.927
RANK 3 vs 4	1.235	0.572	2.668

- Rank 4 category is the base category.
- Having attended an undergraduate institution with a rank of 1, versus an institution with a rank of 4, increases the odds of admission by 4.718.

SAS Probit Output

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Std Est	Wald Error	χ^2	$Pr > \chi^2$
Intercept	1	-3.3225	0.6633	25.0872	< 0.0001
GRE	1	0.00138	0.000650	4.4767	0.0344
GPA	1	0.4777	0.1972	5.8685	0.0154
RANK 1	1	0.9359	0.2453	14.5606	0.0001
RANK 2	1	0.5205	0.2109	6.0904	0.0136
RANK 3	1	0.1237	0.2240	0.3053	0.5806

Probit Interpretation

- The probit regression coefficients give the change in the probit index, also called the Z score, for a unit increase in the regressor.
- For a unit change in GRE, the Z score increases by 0.001.
- For a unit change in GPA, the Z score increases by 0.478.
- The coefficients for rank categories have a different interpretation. Here, having attended an undergraduate institution with a rank of 1, versus an institution with a rank of 4, increases the Z score by 0.936 (Rank 4 is the base category).

THANK YOU