

1 Binary Response Models

1.1 General Textbooks

- Wooldridge, J. (2010). Econometric Analysis of Cross-Section and Panel Data. MIT.
- A. C. Cameron and P. K. Trivedi (2005) Microeconometrics: Methods and Applications, Cambridge University Press
- J. Freidman, T. Hastie, R. Tibshirani (2009). The Elements of Statistical Learning. Springer, publishers.

1.2 Limited Dependent Variables

1.2.1 Observational rules

Assume $\exists -\infty < y^* < \infty$ and:

$$y = \mathbf{1}(\infty < y^* < \infty)y^*$$

Trivially this *observational rule* represents the case with no "limit" on the dependent variable.

Binary Response:

$$y = \mathbf{1}(y^* > \alpha)$$

Multinomial Response:

$$y = \operatorname{argmax}(\mathbf{y}^*)$$

1.3 Revealed vs Stated Preference

RP:

- + Using a random utility maximisation framework, RP theory can extract information from choice behaviour based on discrete information embodied in the chosed alternative within a well defined choice set, together with attributes of the DM and a set of characterstics defining the alternatives.
- Reliable to estimate current market behaviour, but usually insufficient variation across key factors along with colinearity
- Deficient data as number of attributes overserved not measurable

SP:

- + Elicits preferences of consumer prior to choice made.
- = Uncertainty here rooted in DM's ability to accurately report preferences, instead of with the analyst in RP world
- Difficult to design a survey that is not subject to bias

1.4 Binary Response Model

We have a random sample of observations on y_i and x_i , where y_i is a binary variable and x_i is a vector of explanatory variables. $i = 1, \dots, N$.

$$E(y_i = 1|x_i) = Pr(y_i = 1|x_i) = F(x_i'\beta)$$

where F is some monotonically increasing CDF.

Note:-

Here we do not need necessarily to choose a parametric form of $F(\cdot)$
We identify K parameters in β but unlike OLS we cannot separately identify σ

1.4.1 Deriving $\hat{\beta}$ (low emphasis)

We have N independent observations on y_i and x_i .

The probability density of y_i conditional on x_i is:
$$\begin{cases} F(y_i|x_i, \beta) & \text{if } y_i = 1 \\ 1 - F(y_i|x_i, \beta) & \text{if } y_i = 0 \end{cases}$$

The joint distribution of the n data point sequence we observe is:

$$f(y_1, \dots, y_n | x_1, \dots, x_n, \beta) = \prod_{i=1}^n F(y_i|x_i, \beta)^{y_i} (1 - F(y_i|x_i, \beta))^{1-y_i}$$

We can write this equivalently as a likelihood function, depending on β :

$$L(\beta) = \prod_{i=1}^n F(y_i|x_i, \beta)^{y_i} (1 - F(y_i|x_i, \beta))^{1-y_i}$$

If the model is correctly specified then MLE $\hat{\beta}_N$ is consistent, efficient and asymptotically normal.

We work with the log-likelihood function:

$$l(\beta) = \sum_{i=1}^n y_i \log(F(x'_i \beta)) + (1 - y_i) \log(1 - F(x'_i \beta))$$

We then find the critical point:

$$\begin{aligned} \frac{\partial l(\beta)}{\partial \beta} &= \sum_{i=1}^n y_i \frac{f(x'_i \beta)}{F(x'_i \beta)} x_i - (1 - y_i) \frac{f(x'_i \beta)}{1 - F(x'_i \beta)} x_i \\ &= \sum_{i=1}^n \frac{y_i - F(x'_i \beta)}{F(x'_i \beta)(1 - F(x'_i \beta))} f(x'_i \beta) x_i \\ &= \sum_{i=1}^n \tilde{u}_i x_i \quad (*) \end{aligned}$$

where \tilde{u}_i is the *generalised-residual*.

Note:-

Note that $\frac{\partial l(\beta)}{\partial \beta}$ is non linear in β , in general no analytical solution can be found so we resort to numerical methods.

If $l(\beta)$ is (strictly) concave, as in the Probit and Logit, then MLE $\hat{\beta}_N$ is unique.

We can use (*) to test for model misspecification based upon, for example LM tests.

1.4.2 Issues with Binary Response Models

Heterogenous effects in non-linear models:

Potential outcomes in the binary model are given by:

$$Y(x) = \mathbf{1}(\alpha + \beta x + u \geq 0)$$

The effect of a change from x to x' for an individual with error u is:

$$Y(x') - Y(x) = \mathbf{1}(\alpha + \beta x' + u \geq 0) - \mathbf{1}(\alpha + \beta x + u \geq 0)$$

Note:-

This is distinct from a linear model as here partial effects are a function of both x and u . In linear regression u is additive and thus partial effects are independent of u . Now we need to specify (in the parametric case) a distribution for u on top of GM assumptions and beyond iid and first moments. Now need $F()$

Suppose $\beta > 0$ and $x' > x$. Then:

value of u	$Y(x)$	$Y(x')$	$Y(x') - Y(x)$
$-u \leq \alpha + \beta x \leq \alpha + \beta x'$	1	1	0
$-u \geq \alpha + \beta x' \geq \alpha + \beta x$	0	0	0
$\alpha + \beta x \leq -u \leq \alpha + \beta x'$	0	1	1

The effects are either 0 or 1 dependent on the value of u and this differs across individuals. As this u is unobservable this formulation is not yet of immediate use.

1.4.3 The Partial Effects

Consider the partial effect conditional on two values of X , but averaged wrt u .

$$\begin{aligned} E_u[Y(x') - Y(x)] &= E_u[\mathbf{1}(\alpha + \beta x' + u \geq 0)] - E_u[\mathbf{1}(\alpha + \beta x + u \geq 0)] \\ &= Pr(-u \leq \alpha + \beta x') - Pr(-u \leq \alpha + \beta x) \\ &= F(\alpha + \beta x') - F(\alpha + \beta x) \end{aligned}$$

Then $E_u[Y(x') - Y(x)]$ denotes the proportion of units in the population had their outcomes been affected from a change from x to x' i.e. those with $\alpha + \beta x \leq -u \leq \alpha + \beta x'$.

1.4.4 The General Partial Effect over continuous X

For continuous X we can instead write:

$$\frac{\partial E_u[Y(x)]}{\partial x} = \frac{\partial F(\alpha + \beta x)}{\partial x} = \beta f(\alpha + \beta x)$$

In binary response models we can regard PEs as a random variable associated with X .

This means we can compute a PE for each observation in a sample.

In this sense it may be of interest to obtain summary measures of its distribution, like the mean or median.

Definition 1.4.1

Partial Effect of x_k :

Initial value of (x, ε) is (x_0, ε_0) .

The change is $x_k + \Delta_k$, where Δ_k is a vector of zeros except at position k , a 1.

$$PE_k(x_0) = g(x_0 + \Delta_k, \beta) - g(x_0, \beta)$$

Average Partial Effect:

APE_k is $PE_k(x_0)$ averaged over the distribution of observables x .

$$\text{Continuous } X : APE_k = E_x[PE_k(x_0)] = E_X[\beta_k f(x' \beta)]$$

$$\text{Discrete } X : APE_k = E_x[PE_k(x_0)] = E_X[F([x + \Delta_k]' \beta) - F(x' \beta)]$$

Partial Effect at the Average:

$$\begin{aligned} PEA_k &= PE_k(x_0 = E(x)) \\ &= g(E(x) + \Delta_k, \beta) - g(E(x), \beta) \end{aligned}$$

Note:-

PEs in Linear Regression:

In this model with constant linear partial effects:

$$PE_k(x_0) = APE_k = PEA_k = \beta_k$$

But we can introduce non constant PEs through multiplicative effects- interactions between observables.

Definition 1.4.2

The Probit Model:

$$Y_i^* = x_i' \beta + u_i; \quad u_i \sim N(0, \sigma^2)$$

$$Y_i = \mathbf{1}(Y_i^* > 0)$$

We can think of Y_i^* as an index of net utility from an action involving two (or more) choices/states.

$$\begin{aligned} \Rightarrow Pr(Y_i = 1|x_i) &= Pr(Y_i^* > 0|x_i) \\ &= Pr(u_i > -x_i' \beta | x_i) = 1 - F(-x_i' \beta) \\ &= Pr\left(\frac{u_i}{\sigma} > -\frac{x_i' \beta}{\sigma} | x_i\right) = 1 - \Phi\left(-\frac{x_i' \beta}{\sigma}\right) \end{aligned}$$

Thus β is not separately identified to σ , which is thus usually normalised to 1

Example. Common Distributions and their Conditional PEs:

LPM: $F() = I()$, $PE_k = \frac{\partial(x' \beta)}{\partial x_k} = \beta_k$

Probit: $F() = \Phi()$, $PE_k = \beta_k \phi(x' \beta)$

Logit: $F() = \Lambda()$, $PE_k = \beta_k \Lambda(x' \beta) (1 - \Lambda(x' \beta))$

While the LPM may give probability estimates outside the unit interval, and have constant PEs (and het by construction). If we only care about PEs for the average individual then this may not matter, especially in large samples.

1.4.5 Sparse Covariates

With no covariates or sparse and discrete covariates, linear models and the associated estimation techniques (e.g. 2SLS) are no less appropriate for LDVs than for other kinds of DVs. (Angrist JBES 2001)

1.4.6 Unobserved Heterogeneity

Proposition 1.4.1. In probit models, neglected (unobserved) heterogeneity is a much more serious problem than in linear models as even if it is independent of X , probit coefficients are inconsistent.

We will show that this statement is only **partially** true.

Proof. Probit Parameters become inconsistent:

We illustrate with a one period cross sectional probit model:

$$Pr(y_i = 1 | x_i, c_i) = \Phi(x_i' \beta + c_i)$$

where x_i is $k \times 1$, $x_{1i}/equiv 1$ and c_i is an unobserved scalar, denoting the unobserved heterogeneity

Our parameter of interest is the partial effect estimation of a given x_j on $Pr(y_i = 1 | x_i, c_i)$

Consider the standard latent variable model:

$$y_i^* = x_i' \beta + \gamma c_i + \varepsilon_i$$

$$y_i = \mathbf{1}(y_i^* > 0)$$

$$\varepsilon_i | x_i, c_i \sim N(0, 1)$$

$$c_i | x_i \sim N(0, \tau^2)$$

c_i normalised such that $E(c) = 0$ and assumed independent of x .

$$\gamma c_i + \varepsilon_i | x_i \sim N(0, \gamma^2 \tau^2 + 1)$$

$$\begin{aligned} \Rightarrow Pr(y_i = 1 | x_i) &= Pr(\gamma c_i + \varepsilon_i > -x_i' \beta | x_i) \\ &= \Phi(-x_i' \beta / \sigma) \end{aligned}$$

where $\sigma^2 = \gamma^2 \tau^2 + 1$

Thus we see that this probit model of y on x consistently estimates β / σ

Since $\sigma = (\gamma^2\tau^2 + 1)^{1/2} > 1$ then $|\beta_j/\sigma| < |\beta_j|$ we have **attenuation bias** in estimated parameters. Attempting to normalise $\gamma^2\tau^2 + 1 = 1$ would imply $\gamma^2\tau^2 = 0$ and so there is no unobserved heterogeneity. □

Proof. Unobserved heterogeneity does not affect partial effect estimates.

For given values of x_i and c_i the partial effect of interest is:

$$\frac{\partial Pr(y_i = 1|x_i, c_i)}{\partial x_{ij}} = \beta_j \phi(x'_i \beta + \gamma c_i)$$

Given $E(c) = 0$, evaluating at $c = 0$:

Ground truth: $\beta_j \phi(x'_i \beta)$

Our estimate: $(\beta_j/\sigma) \phi(x'_i \beta \sigma)$

Thus probit does not give correct partial effects for $c = 0$

Estimating APE by averaging over c:

$$E_c \left[\frac{\partial Pr(y_i = 1|x_i, c_i)}{\partial x_{ij}} \right] = E_c [\beta_j \phi(x'_i \beta + \gamma c_i)]$$

But:

$$\begin{aligned} E_c \left[\frac{\partial Pr(y_i = 1|x_i, c_i)}{\partial x_{ij}} \right] &= \frac{\partial}{\partial x_{ij}} E_c [Pr(y_i = 1|x_i, c_i)] \\ &= \frac{\partial}{\partial x_{ij}} E_c [E_{y|x,c}[y_i = 1|x_i, c_i]] \\ &= \frac{\partial}{\partial x_{ij}} E_{y|x}[y_i|x_i] \\ &= \frac{\partial}{\partial x_{ij}} Pr(y_i = 1|x_i) \\ &= \frac{\partial}{\partial x_{ij}} Pr(\gamma c_i + u_i > -x'_i \beta) \\ &= \frac{\partial}{\partial x_{ij}} \Phi(-x'_i \beta / \sigma) \\ &= \beta_j / \sigma \phi(x'_i \beta / \sigma) \end{aligned}$$

For this measurement, the omitted heterogeneity is not a problem when independent of x . We can consistently estimate the APEs, conditional on normality of c and the probit model. The reason why the first PE was inconsistent was due to conditioning on the unobservable c_i , which we cannot do when estimating β .

But we can avoid this issue in estimating APEs by exploiting the law of iterated expectations to removing conditioning on c_i . □

Omitted heterogeneity in linear models is not a problem when independent of x . (subsumed by error term). Omitted heterogeneity in probit models also not a problem when independent of x . As ignoring it consistently estimates APEs.

Note:-

If c is correlated with x or otherwise dependent on x (i.e. $Var(c|x)$ is a function of x), then omission of c will mean we cannot get consistent APE estimates.