Econometrics 6027
Lecture 2
Binary Choice Models and Multiple Discrete Choice Models

1 Classes of Discrete Variable

There are two classes of discrete variables:

1. Binary
   A variable Y_t which can take a value of either 1 or 0

2. Multinomial
   Which can be further classified as:

- Categorical:

  $y = 1$ if income $< \$10,000$

  $y = 2$ if $\$10,000 \leq$ income $\leq \$20,000$

  $y = 3$ if $\$20,000 <$ income

  A further classification of categorical variables depends on whether the specific outcomes taken by that variable have a natural ordering or sequence.

  – Nominal/unordered categorical

    $y = 1$ if the mode of transport is by car
    $y = 2$ if the mode of transport is by bus
    $y = 3$ if the mode of transport is by train

  – Sequential

    $y = 1$ if an individual is working
    $y = 2$ if an individual is working part-time

- Non-categorical

  $y =$ the number of TV sets in a household

The characteristics of any discrete variable dictate the methods available for model solution.

2 About Binary Choice Models

Theoretical framework: Consider a binary dependent variable y which has only two possible outcomes (0 and 1), and a vector of explanatory variables x thought to influence the realization of y.

The unconditional expectation of the binary variable y is by definition a probability:

$$E(y) = \Pr(y = 1)$$

Further, let the set of explanatory variables x influence the outcome of y. Then, the conditional expectation of y given x is:

$$E(y|x) = \Pr(y = 1|x)$$

Relate this term to the standard regression analysis:

$$y = F(x, \beta) + u$$

has the conditional expectation

$$E(y|x) = F(x, \beta) + E(u|x) = F(x, \beta)$$

- Therefore, the standard regression functional $F(x, \beta)$ is a representation of the conditional expectation of y given x.
- If the dependent variable in a regression relationship is binary, then the regression function equates directly to the conditional probability of observing $y = 1$.
- Thus, the characteristics of binary choice models depend on the way we specify $F(x, \beta)$.

The latent variable approach: Assume that there is some underlying (and unobserved) latent *propensity* variable y^* where

$$y^* \in (-\infty, \infty).$$

We do not observe y^* directly, but we do observe a binary outcome y such that

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

where $\mathbf{1}(.)$ is the *indicator function* taking the value 1 if the condition within parentheses is satisfied, and 0 otherwise.

Define the latent equation in linear form,

$$y^* = x'\beta + u,$$

where u is random with symmetric density $f(.)$ and corresponding cdf $F(.)$.

We now have that

$$\begin{aligned} E(y|x) &= \Pr(y = 1|x) \\ &= \Pr(y^* > 0|x) \\ &= \Pr((x'\beta + u) > 0) \\ &= \Pr(u > -x'\beta) = F(x'\beta). \end{aligned}$$

By specifying an appropriate distribution function for u, we can derive binary choice models.

Example Let y be a labour force participation variable with $y = 1$ if the individual works and $y = 0$ otherwise and let the outcome (working, non-working) being described by the state-specific utilities $U^*(y)$, with

$$\begin{aligned} U^*(y=1) &= x'\beta_1 + u_1 \\ U^*(y=0) &= x'\beta_0 + u_0 \end{aligned}$$

Participation in the labour force requires that $U^*(y=1) > U^*(y=0)$, such that

$$\begin{aligned} y &= 1[U^*(y=1) > U^*(y=0)] \\ &= 1(x'\beta_1 + u_1 > x'\beta_0 + u_0) \\ &= 1[u_1 - u_0 > -x'(\beta_1 - \beta_0)]. \end{aligned}$$

We cannot identify both parameters $\beta_1$and β_0, but we can identify the difference $\beta_1 - \beta_0$ Hence,

$$y = 1(y^* > 0)$$

where

$$y^* = x'(\beta_1 - \beta_0) + (u_1 - u_0) = x'\beta + u.$$

Thus, the latent variable approach to a binary choice model can be derived from an economic model of behavior.

3 The Linear Probability Model (LPM)

Consider a binary dependent variable y and a (kx1) vector of explanatory variables x. We may specify the conditional probability directly as:

$$\Pr(y = 1|x) = F(x, \beta) = x'\beta;$$

Introducing random disturbances, we have

$$y = x'\beta + u,$$

where u represents the stochastic disturbance term in the relationship, $f(u)$ represents its density and $E(u|x) = 0$ by definition. This model is known as the Linear Probability Model (LPM).

For a sample of n observations $\{y_i, x_i\}$ drawn at random from a population,

$$y_i = x_i'\beta + u_i.$$

The Ordinary Least Squares estimation procedures may be applied.

The LPM might therefore be considered a first-order approximation to the arbitrary nonlinear probability function $F(x_i, \beta)$; that is,

$$F(x, \beta) \approx F(x_0, \beta) + (x - x_0)' \frac{\partial F(x_0 \beta)}{\partial \beta} = x'\beta$$

using a first-order Taylor series expansion around $x = x_0$.

Problems with the LPM:

- disturbance terms are non-normal;

$$\begin{aligned} u_i &= 1 - x_i'\beta \text{ with probability } f(u_i) = x_i'\beta \text{ (for } y = 1) \\ u_i &= -x_i'\beta \text{ with probability } f(u_i) = 1 - x_i'\beta \text{ (for } y = 0) \end{aligned}$$

- disturbance terms are heteroskedastic;

$$\begin{aligned} var(u_i|x_i) &= E(u_i^2|x_i) \\ &= (-x_i'\beta)^2(1 - x_i'\beta) + (1 - x_i'\beta)^2(x_i'\beta) \\ &= (x_i'\beta)(1 - x_i'\beta) \\ &= \Pr(y_i = 1|x_i)\Pr(y_i = 0|x_i). \end{aligned}$$

- the conditional expectation is not bounded between zero and one;

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

  Instead, it is defined over the entire real line. What is the expected value of X_i here? It is $P(Y = 1|X)$. This is a probability, and so the support (the domain) of $X_i'\beta$ is not necessarily between zero and one. However, the support of Y has to be between zero and one by definition.

  There are thus two problems: firstly, in linear probability models, the predicted Y's can be > 1, so that when you build weights to address heteroskedasticity (as we will see in the next section), they can be $\sqrt{-1}$, which doesn't exist! Secondly, after they have been weighted, the predicted Y's can be > 1, which also makes no sense.

Weighted Least Squares: A Solution We transform this model to give it a constant variance using a $\frac{1}{W_i}$ weight.

$$\frac{Y_i}{W_i} = \frac{X_i}{W_i}\beta + \frac{u_i}{w_i}$$

In order to obtain a constant $\frac{u_i}{w_i}$, we use the following weight (with $var(Y_i) = 1$):

$$w_i = \sqrt{(x_i'\widehat{\beta})(1 - x_i'\widehat{\beta})}$$

calculated from a first-stage estimation. The weights are a function of our β since we use a $\hat{\beta}$ in the estimation of the weights, so we introduce some randomness, but it may nonetheless look homoskedastic. We need a consistent estimator of the β: if we estimate with OLS, it will be consistent but heteroskedastic. If we use w_i, the adjusted model then becomes

$$\frac{y_i}{w_i} = \left(\frac{x_i}{w_i}\right)'\beta + \frac{u_i}{w_i}$$

Which will generate $\hat{\beta}_{LPM}$. We can draw inference with this new OLS model, where we've corrected the errors.

$$var(\frac{u_i}{x}) = \frac{var(u_i)}{x^2}$$

However, this model still does not return probabilities within the range $[0, 1]$. We must still assume that this is the case, or impose the restriction on the data. A better solution is to re-specify, or transform the regression model itself to constrain the probability outcome.

LPM Marginal Effects The change in the expected value of y from x is:

$$\frac{\partial E(y_i|x_i)}{\partial x_i} = \beta$$

since:

$$E(y = 1|x_i) = Pr(y_i = 1|x_i)$$

4 The Probit Model

Any probability function is non-linear. If, however, we assume that $u_i \sim N(0, \sigma)$, then $Pr(y_i = 1|x_i)$ should be normal.

$$y_i = Pr(y_i = 1|x_i) + u_i$$

But if $Pr(y_i = 1|x_i)$ is normal, we cannot represent it linearly. If it is normal, the distribution (PDF) is:

$$f(u_i)\phi(u_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-u^2}{2\sigma^2}}$$

and the CDF is:

$$F(x_i, \beta) = \Phi(z) = \int_{-\infty}^{z} \phi(u) du$$

Since the normal is nonlinear, this integral is also non-linear and we have:

$$Pr(y_i = 1|x_i) = \Phi(\frac{x_i'\beta}{\sigma})$$

Where the CDF is $\frac{x_i'\beta}{\sigma}$. This is obtained by writing $P(y_i = 1|x_i)$ for u. If looking at $y_i = u - \frac{x_i'\beta}{\sigma}$. The probability of $u_i = 1 - x_i'\beta$ is $x_i'\beta$. If non-linear, the probability will be $F(x_i'\beta)$. When u is assumed normally distributed, parameters must scaled to force the variance of u to unity. That's why we replace $\Phi(x_i'\beta)$ then normalize by dividing by σ.

$$\begin{aligned}\Pr(y=1|x) &= \Pr(u > -x'\beta) \\ &= \Pr[\frac{u}{\sigma} > -x'\frac{\beta}{\sigma}] \\ &= \Pr[z > -x'\frac{\beta}{\sigma}] \\ &= \Phi(x'\frac{\beta}{\sigma}) \\ y_i &= \Phi(\frac{x_i'\beta}{\sigma}) + u_i.\end{aligned}$$

Note: The function $\Phi(z)$ is a monotone increasing function of z. Moreover, the model returns well-defined probabilities:

$$\begin{aligned}F(x_i,\beta) &\rightarrow 0 \text{ as } x_i^{'}\beta \rightarrow -\infty \\ F(x_i,\beta) &\rightarrow 1 \text{ as } x_i^{'}\beta \rightarrow \infty\end{aligned}$$

Indeed, that's why we use it! However, because the transformed regression function is non-linear in β, we can no longer use OLS and must move to Maximum Likelihood solution techniques.

4.1 Estimation of the Probit Model

We construct a likelihood estimator with parameters β, σ.

Consider a sample of n observations $\{y_i, x_i\}$, where y_i is binary. Assume $y_i = 1(y_i^* > 0)$ and $1 - y_i = 1(y_i^* \leq 0)$ for $y_i^* = x'\beta + u_i$.

For any vector β, the probability of observing y_i conditional on x_i for an individual is

$$L_i(\beta,\sigma|x_i,y_i) = Pr(y_i = 1|x_i) * Pr(y_i = 0|x_i)$$

and for all individuals it is

$$L^T(\beta,\sigma|x,y) = \prod_{i=1}^{m} L_i = \prod_{i=1}^{m}[Pr(y_i = 1|x_i)^{y_i} * Pr(y_i = 0|x_i)^{1-y_i}]$$

Since for the Probit model,

$$\begin{aligned}\Pr(y_i &= 1|x_i;\beta) = \Phi(x_i'\beta), \\ \Pr(y_i &= 0|x_i;\beta) = 1 - \Phi(x_i'\beta).\end{aligned}$$

and normalizing by dividing by the variance, we have:

$$L^T(\beta,\sigma|x,y) = \prod_{i=1}^{m}[\Phi(\frac{x_i'\beta}{\sigma})^{y_i} * (1 - \Phi(\frac{x_i'\beta}{\sigma}))^{1-y_i}]$$

Taking logs simplifies the equation:

$$\ln L^T(\beta,\sigma|x_i,y_i) = \sum_{i=1}^{m}\Big[y_i \ln\Phi(\frac{x_i'\beta}{\sigma}) + (1-y_i)\ln(1-\Phi(\frac{x_i'\beta}{\sigma}))\Big]$$

since y_i can only take on values of 0 and 1, half of the terms in the summation are equal to zero because of the indicator function.

First Order conditions

$$\beta,\sigma \arg\max_{\beta,\sigma} lnL^T$$

The first order condition for β (also known as the score function) is non-linear:

$$\frac{\partial \ln L^T}{\partial\beta} = \sum_{i=1}^{n}\Big[y_i\frac{\phi(\frac{x_i'\beta}{\sigma})}{\Phi(\frac{x_i'\beta}{\sigma})}x_i + (1-y_i)\frac{\phi(\frac{x_i'\beta}{\sigma})\frac{x_i}{\sigma}}{1-\Phi(\frac{x_i'\beta}{\sigma})}\Big] = 0$$

where Φ is the CDF, the cumulative distribution function and ϕ is the PDF, the probability distribution function (the derivative of the CDF).

The first order condition for σ is:

$$\frac{\partial lnL^T}{\partial\sigma} = \sum_{i=1}^{n}[Y_i\frac{\phi(\frac{X_i'\beta}{\sigma})}{\Phi(\frac{X_i'\beta}{\sigma})} + (1-Y_i)\frac{\phi(\frac{X_i'\beta}{\sigma})(\frac{-X_i'\beta}{\sigma^2})}{1-\Phi(\frac{X_i'\beta}{\sigma})}] = 0$$

To find $\hat{\beta}$ and $\hat{\sigma}$ requires solving a system of two equations in two unknowns, which is not straightforward to solve. You can start with an initial value, and work on a gradient until get a first order condition (FOC) value close to zero, which will give you your optimal β.

Probit Marginal Effects

$$\begin{aligned}
\frac{\partial E(Y_i=1|X_i)}{\partial X} &= \frac{\partial P(Y_i=1|\phi)}{\partial X} \\
&= \phi(\frac{X_i\beta}{\sigma})^2 \\
\frac{\partial E(Y_i|X)}{\partial X} &= \frac{\partial Pr(Y_i=1|X)}{\sigma X} \\
&= \frac{\partial\Phi(\frac{X_i'\beta}{\sigma})}{\partial X'} \\
&= \phi(\frac{X_i'\beta}{\sigma})\frac{\beta_p}{\sigma}
\end{aligned}$$

But this depends on i, so each individual would have a different marginal effect. If we have the normal distribution, $\sigma = 1$, the denominators will disappear and we will have β_p. Let's assume $\sigma = 1$. Then,

$$\phi(X_i'\beta)\beta_p = \beta_{LPM}$$

where $\phi(X_i\beta)$ is between zero and one.

5 The Logit Model

If u_i is not normal, we can assume that it follows the logistic (Λ) distribution. In this case,

$$F(x_i, \beta) = \Lambda(x_i^{'}\beta)$$

where

$$\Lambda(x_i'\beta) = \Pr(y_i = 1|x_i; \beta) = \frac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)} = \frac{e^{X'\beta}}{1 + e^{X\beta}} = \frac{\exp(z)}{1 + \exp(z)} = \frac{1}{1 + \exp(-z)}$$

is the CDF of the Logistic function with $u \sim \mathit{Logistic}$ and a mean of zero $E(u_i|X_i) = 0$.

Like the Probit model, the Logit CDF is a monotone increasing function of z that returns well-defined probabilities between 0 and 1 but is non-linear in β and which must thus be estimated by Maximum Likelihood solution techniques.

5.1 Estimation of the Logit Model

We construct a maximum likelihood estimator for the logit model.

Since

$$\begin{aligned}
\Pr(y_i &= 1|x_i; \beta) = \Lambda(x_i'\beta) = \frac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)}; \\
\Pr(y_i &= 0|x_i; \beta) = 1 - \Lambda(x_i'\beta) = \frac{1}{1 + \exp(x_i'\beta)},
\end{aligned}$$

The likelihood for a given individual will be:

$$\begin{aligned}
L_i &= \Lambda(x_i'\beta)^{y_i}(1 - \Lambda(x_i'\beta))^{1-y_i} \\
&= (\frac{e^{x'\beta}}{1 + e^{x'\beta}})^{y_i}(1 - \frac{e^{x'\beta}}{1 + e^{x'\beta}})^{1-y_i} \\
&= (\frac{e^{x'\beta}}{1 + e^{x'\beta}})^{y_i}(\frac{1}{1 + e^{x'\beta}})^{1-y_i} \\
&= (e^{x_i\beta})^{y_i}\frac{1}{1 + e^{x'\beta}} \\
L^T &= \prod_{i=1}^{n}(e^{x'\beta})^{y_i}\frac{1}{1 + e^{x'\beta}} \\
lnL^T &= \sum_{i=1}^{n}[y_i(x_i\beta) - ln(1 + e^{x_i'\beta})]
\end{aligned}$$

This is non-linear in β, but easier to estimate than probit and LPM.

First Order conditions The FOCs/ score function is:

$$\frac{\partial lnL^T}{\partial \beta} = \sum_{i=1}^{n}(y_i x_i' - \frac{e^{x'\beta}}{1+e^{x\beta}}x_i) = \sum_{i=1}^{n}(y_i - \frac{e^{x'\beta}}{1+e^{x\beta}})x_i = 0$$

which gives β logistic, $\hat{\beta}_L$.

Logit Marginal Effects

$$\hat{\beta}_L = \frac{\partial \Lambda(x_i'\beta)}{\partial x} = \frac{\partial \Pr(y_i = 1|x;_i\beta)}{\partial x_{ij}} = \frac{\exp(x_i'\beta)}{[1+\exp(x_i'\beta)]^2}\beta_j$$

6 Comparing LPM, Probit and Logit Models

6.1 Interpreting/comparing binary choice model coefficients

The coefficients are:

$$\begin{aligned} LPM &: \quad \Pr(y_i = 1|x_i;\beta) = x_i'\beta \\ Probit &: \quad \Pr(y_i = 1|x_i;\beta) = \Phi(x_i'\beta) \\ Logit &: \quad \Pr(y_i = 1|x_i;\beta) = \Lambda(x_i'\beta) \end{aligned}$$

If the parameter β_j associated to the $j-$th explanatory variable is positive (negative), then

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

will increase (decrease) with an increase in x_j.

The marginal effects are:

$$\begin{aligned} LPM &: \quad \frac{\partial \Pr(y_i = 1|x_i;\beta)}{\partial x_{ij}} = \beta_j \\ Probit &: \quad \frac{\partial \Pr(y_i = 1|x_i;\beta)}{\partial x_{ij}} = \phi(x_i'\beta)\beta_j \\ Logit &: \quad \frac{\partial \Pr(y_i = 1|x_i;\beta)}{\partial x_{ij}} = \frac{\exp(x_i'\beta)}{[1+\exp(x_i'\beta)]^2}\beta_j \end{aligned}$$

Coefficients are not the same as marginal effects. By implication, slope estimates are **not** directly comparable amongst models (eg. variance of disturbances in Logit model and the Probit model are different). Hence, the parameters are scaled differently as well.

Notice also that:

- the marginal effects in the LPM are constant (ie. independent of the data)
- the marginal effects in the Probit and Logit models depend on x_i.

6.2 Comparing LPM and Probit Estimators

The probit and LPM estimator should not be equal, they should be different. One is linear, one is not; they may both be consistent. Note that the probit functional form has to be between zero and one by definition and $\hat{\beta}_{LPM}$ has to be, and should be, between zero and one, either by assumption or by adjustment to get $0 < \hat{\beta}_{LPM} < 1$.

If we know that the estimators should be different, we can find the ratio ($\hat{\beta}_{LPM}$ to $\hat{\beta}_{P}$.) between the two by linearizing the β.

$$\begin{aligned}
\hat{Y}_i &= X_i'\hat{\beta}_{LPM} + \hat{u}_i \\
\hat{Y}_i &= F(X_i\hat{\beta}_P) + \hat{u}_i \\
X_i'\hat{\beta}_{LPM} &= F(X_i\hat{\beta}_p)
\end{aligned}$$

Yet we know from each model's marginal effects that (when $\sigma = 1$):

$$\begin{aligned}
\frac{\partial E(y_i = 1|x_i)}{\partial x} &= \hat{\beta}_{LPM} \\
\frac{\partial E(y_i = 1|x_i)}{\partial x} &= \phi(x_i'\beta)\beta_p \\
\phi(X_i'\beta)\beta_p &= \hat{\beta}_{LPM} \\
\beta_p &= \frac{\hat{\beta}_{LPM}}{\phi(x_i'\beta)}
\end{aligned}$$

and since $\phi(x_i'\beta)$ is between zero and one, we can see that $\hat{\beta}_p$ should be bigger than $\hat{\beta}_{LPM}$. Or, equivalently, we can say that $\hat{\beta}_{LPM}$ should be smaller than $\hat{\beta}_p$. In particular, the ratio should be:

$$\hat{\beta}_{LPM} \approx 0.25\hat{\beta}_p \qquad \hat{\beta}_p \approx 0.625\hat{\beta}_{LPM}$$

If you get a ratio different from this, say $\hat{\beta}_p = 1, \hat{\beta}_{LPM} = 0.7$, either the LPM estimator is not consistent, or the normality assumption on the error is incorrect. To see if the latter is the case, we can test for the probit model's normality of errors. If errors are not normal, β_p is not consistent and it is useless. But this doesn't mean that β_{LPM} is consistent. The errors may be non-normal AND β_{LPM} may be inconsistent. If the ratio is preserved, can see that the model is good.

6.3 Comparing LPM, Probit and Logit Estimators

We have:

Linear: $Y_i = X_i\beta + u_i$

Probit: $Y_i = \Phi(X_i\beta) + u_i$

Logit: $Y_i = \Lambda(X_i\beta) + u_i$

If all three estimators are consistent, then the marginal effect, or change in $E(Y_i)$, should be the same. The marginal effects for each estimator are:

$$\hat{\beta}_{LPM} = \frac{\partial E(u_i)}{\partial X} = \frac{\partial ln(Y_i - 1|X)}{\partial X_i} = \lambda(X_i'\beta)$$

$$\hat{\beta}_P = \phi(X_i\beta)$$

$$\hat{\beta}_L = \frac{\partial \Lambda(X_i'\beta)}{\partial X}$$

If the marginal effect of each estimator is the same, then we have:

$$\frac{\partial E(u_i)}{\partial X} = \lambda(X_i'\beta) = \frac{\partial \Lambda(X_i'\beta)}{\partial X}$$

Yet we know that the derivative of $\Lambda(X'\beta)$ with respect to $X'\beta$ or, equivalently, u, is just the PDF of $\Lambda(X'\beta)$. We know that the CDF of the logit is $\Lambda(u) = \frac{e^u}{1+e^u}$. So to find that PDF of the logit, we take the derivative of $\Lambda(u)$.

$$\begin{aligned}
\lambda(u) &= \frac{\partial \Lambda(u)}{\partial u} \\
&= \frac{\partial \frac{e^u}{1+e^u}}{\partial u} \\
&= \frac{e^u(1+e^u) - e^u e^u}{(1+e^u)^2} \\
&= \frac{e^u + (e^u)^2 - (e^u)^2}{(1+e^u)^2} \\
&= (\frac{e^u}{1+e^u})(\frac{1}{1+e^u}) \\
&= \Lambda(u)(1-\Lambda(u))
\end{aligned}$$

So, in words, the PDF of the logistic equals the CDF times one minus the CDF. Thus the marginal effects for each model are:

$$\begin{aligned}
LPM &: \quad \frac{\partial \Pr(y_i = 1|x_i;\beta)}{\partial x_{ij}} = \beta_j \\
Probit &: \quad \frac{\partial \Pr(y_i = 1|x_i;\beta)}{\partial x_{ij}} = \phi(x_i'\beta)\beta_j \\
Logit &: \quad \frac{\partial \Pr(y_i = 1|x_i;\beta)}{\partial x_{ij}} = \hat{\beta}_L \Lambda(x'\beta)(1-\Lambda x'\beta) = \hat{\beta}_L \frac{e^u}{1+e^u}(1 - \frac{e^u}{1+e^u})
\end{aligned}$$

If all three models are consistently estimated, β_{logit} will be the largest. The ratios should be:

$$LPM : 0.7 \quad Probit : 1 \quad Logit : 1.4$$

If ratio is not preserved between 0.7 and 1: if, for example, have LPM: 0.7, Probit: 1, Logit:1/0.0625; then LPM is not consistent. Probit and logit are consistent, can use probit. Probit is more efficient because the tails are smaller. If all the ratios are not preserved, then we don't know. Remain with logistic.

6.4 Empirical Example: childcare take-up estimates

(y=1 if the women uses paid childcare, y=0 otherwise)

		Parameter	Estimates
Variable	LPM	Probit	Logit
single woman	-0.059	-0.184	-0.310
other children aged 5+	-0.101	-0.318	-0.540
woman works	0.152	0.430	0.713
left school at 18	0.109	0.310	0.520
attended college/university	0.160	0.458	0.757
youngest child aged 2	0.186	0.556	0.928
youngest child aged 3-4	0.309	0.882	1.458
receives maintenance	0.089	0.264	0.432
constant	0.153	-0.995	-1.645

The reference in all cases is:

- a married women who
- doesn't work
- has left school at 16
- has one child aged less than 2 and
- receives no maintenance.

For the reference household, all explanatory variables take a value of 0, which leads to probability estimates in each model of:

LPM: $\Pr(y_i = 1|x_i) = x_i'\widetilde{\beta} = 0.153$

Probit: $\Pr(y_i = 1|x_i) = \Phi x_i'\widetilde{\beta} = \Phi(-0.995) = 0.161$

Logit: $\Pr(y_i = 1|x_i) = \Lambda x_i'\widetilde{\beta} = \exp(-1.645)/[1 + \exp(-1.645)] = 0.162$

How, for example, does the probability change for women who attended university?

LPM: $\Pr(y_i = 1|x_i) = x_i'\widetilde{\beta} = 0.153 + 0.160 = 0.313$

Probit: $\Pr(y_i = 1|x_i) = \Phi x_i'\widetilde{\beta} = \Phi(-0.995 + 0.458) = \Phi(-0.537) = 0.296$

Logit: $\Pr(y_i = 1|x_i) = \Lambda x_i'\widetilde{\beta} = \frac{\exp(-1.645+0.757)}{1+\exp(-1.645+0.757)} = 0.291$

7 Statistical Inference in Binary Choice Models

For the LPM, estimated standard errors can be derived easily. Don't forget LPM is heteroskedastic.

For the Probit and Logit models,

$$\sqrt{n}(\widetilde{\beta} - \beta) \overset{a}{\sim} N(0, I(\widetilde{\beta})^{-1})$$

Computer software for ML estimation evaluates the variance-covariance matrix $V(\widetilde{\beta})$ directly.

Hence, statistical inference and hypothesis testing can then be carried out using standard inferential techniques.

Measures of goodness-of-fit In order to assess the accuracy with which a binary choice model approximates the observed data, 2 measures based on likelihood ratios are proposed and are attributed to Cragg and Uhler (1970) and to McFadden (1974) .

Let L_U - represent the likelihood for the full unrestricted model

Let L_R - represent the likelihood for a restricted model estimated on an intercept alone.

Then the formulation for two proposed measures are as follows:

Cragg and Uhler: $pseudo - R^2 = \frac{L_U^{2/n} - L_R^{2/n}}{1 - L_R^{2/n}}$

McFadden: $pseudo - R^2 = 1 - \frac{\ln L_U}{\ln L_R}$

An alternative outcome-based measure of performance or fit in binary choice models evaluates the *proportion of correct predictions.*

Let $\widetilde{P}_i = \Pr(y_i = 1|x_i) = \Phi(x_i'\widetilde{\beta})$ for a Probit model

Use the following rule to predict states for each observation

$$\widetilde{y}_i = 1\left(\widetilde{P}_i > 0.5\right)$$

The proportion of correct predictions may then be defined as:

$$P = \frac{1}{n}\sum_{i=1}^{n} 1\left(y_i = \widetilde{y}_i\right)$$

Testing the overall significance of the regression Let L_U- represent the likelihood for the full model

Let L_R- represent the likelihood for a restricted model

$r =$ the number of restrictions imposed

To test the joint significance of the slope parameters in the ML model of binary choice use the following statistic

$$-2\ln\left(\frac{L_R}{L_U}\right) = 2\left(\ln L_U - \ln L_R\right) \sim \chi_r^2$$

For example,

$$\begin{aligned} H_0 &: \beta_2 = \beta_3 = ... = \beta_k = 0 \\ H_1 &: \text{ at least one } \beta_j \neq 0,\ j = 2, ..., k \end{aligned}$$

The technique can be generalized to test for a subset of restrictions.

8 Using Binary Choice Models

8.1 Simulating transitions using binary choice models

Empirical binary choice models can be used to simulate individual responses to exogenous shocks. Consider a model of labour market participation which includes the level of non-means tested benefit income among the exogenous variables describing the decision to work, and suppose that a policy option is being considered which would abolish the non-means tested benefits. How might such a policy impact on labour market decisions?

Let

$$y_i^* = x_i'\beta + u_i,$$

consider now an exogenous shock on the exogenous variables x_i. Thus, x_i becomes $x_i^N \Rightarrow y_i^*$ becomes y_i^N

$$y_i^N = \left(x_i^N\right)'\beta + u_i.$$

How does this exogenous shock impact on the probability $P_{i(j\to k)}$ of transition from any state j to k, $(j,k=0,1)$?

A correct way to find transitions probabilities is found in Duncan and Weeks (1970) . They represent the transition probabilities as

$$\begin{aligned}
P_{i(0\to 0)} &= \min\left\{\widehat{p}_{i0}^B, \widehat{p}_{i0}^N\right\} \\
P_{i(0\to 1)} &= 1\left[\widehat{p}_{i0}^B > \widehat{p}_{i0}^N\right]\left(\widehat{p}_{i0}^B - \widehat{p}_{i0}^N\right) \\
P_{i(1\to 0)} &= 1\left[\widehat{p}_{i0}^B < \widehat{p}_{i0}^N\right]\left(\widehat{p}_{i0}^N - \widehat{p}_{i0}^B\right) \\
P_{i(1\to 1)} &= \min\left\{\widehat{p}_{i1}^B, \widehat{p}_{i1}^N\right\}
\end{aligned}$$

In other words, one needs only to difference the two state probabilities for a correct measure of the probability of transition.

8.2 Exogenous & Dummy Regressors Case

Data Generating Process for the Probit Estimator.
Use the same data generating process
The experimental design consist of the following model

$$Y_{it} = \alpha + X_{it}\beta + D_{it}\delta + \varepsilon_{it},$$

where

$$\begin{aligned}
X_{it} &\sim N(0,1) \\
\varepsilon_{it} &\sim N(0,1) \\
D_{it} &= \mathbf{1}\left[(X_{it} + \xi_{it}) > 0\right], \quad \xi_{it} \sim N(0,1) \\
Y_{it} &= \mathbf{1}\left[Y_{it} > 0\right].
\end{aligned}$$

Do the estimation of the model by maximizing the unconditional log-likelihood

$$L_U = \sum_t \{y_{it} \ln \Phi(\alpha + x_{it}\beta + D_{it}\delta) + (1 - y_{it}) \ln \bar{\Phi}(\alpha + x_{it}\beta + D_{it}\delta)\},$$

where $\bar{\Phi}(\alpha+x_{it}\beta+D_{it}\delta) = 1-\Phi(\alpha+x_{it}\beta+D_{it}\delta)$. An estimate for $\theta = \{\alpha, \beta, \delta\}$, is obtained as

$$\widehat{\theta}_{L_U} = \arg\max_{\theta} \left(\frac{\sum_i L_U}{N} \right).$$