EC968
Panel Data Methods

Part 1: Panel Data Analysis

Lecture 4

Mark Bryan
ISER

Lecture 4: Models for discrete variables

- Types of discreteness
- Linear regression
- Latent linear regression
- Binary models: conditional (fixed-effects) logit and random-effects probit and logit

Discreteness

Inherent discreteness involves transitions between states

(*e.g.* employment & unemployment, married, unmarried)

Observational discreteness is an artefact of the observation process

(*e.g.* income questions based on ranges, Likert attitudinal questions)

[Note discrete variables are also sometimes called limited dependent variables]

Forms of discreteness

Censoring/corner solutions generate variables which are mixed discrete/continuous

(*e.g*. hours of work are 0 for non-employed, any positive value for employees)

Truncation involves discarding part of the population

(*e.g*. low-income targeted samples, or earnings models for employees only)

Count variables are the outcome of some counting process

(*e.g*. the number of durables owned, or the number of employees of a firm)

Binary variables reflect a distinction between two states

(*e.g*. unemployed or not, married or not)

Ordinal variables are ordered variables, possibly taking more than two values

(*e.g*. happiness on a scale 1=miserable … 5=ecstatic)

Unordered variables reflect outcomes which are discrete but with no natural ordering

(*e.g*. choice of occupation)

Binary models (1)

Dependent variable is

$$y_{it} = 0 \text{ or } 1$$

Outcome is one of 2 alternative states.

We will focus on modelling this period's state (0 or 1):

- as a function of exogenous explanatory variables and an individual effect (static model)
- as a function of exogenous explanatory variables, an individual effect and last period's state (dynamic model). This allows for *state dependence*. (Note that in this course, we do not consider the effect of time spent in previous states. This *duration dependence* is usually handled by survival analysis techniques.)

Binary models (2)

- There are various possible ways to model binary outcomes:
  - Linear probability model (LPM): very simple (naïve) approach. But does not fit the data well, and not recommended, though may be OK for exploratory analysis.
  - Probit models
  - Logit (logistic) models
- We face similar considerations as for linear models about whether to specify random or fixed effects. But:
  - FE is more difficult: only feasible with logit and cannot easily predict probabilities or calculate marginal effects. "Correlated RE" may be alternative (see later).
  - Note we can also estimate pooled models, correcting SEs for serial correlation, but this is less efficient than modelling individual effects properly (and biased if individual effects are correlated with regressors).

Why are special methods needed ?

Consider the binary variable, $y_{it} = 0$ or 1

Notice that $E(y_{it}) = \Pr(y_{it} = 1).(1) + \Pr(y_{it} = 0).(0) = \Pr(y_{it} = 1)$

where $\Pr(y_{it} = 1)$ is the probability that $y_{it} = 1$

This suggests that a simple way to model y_{it} is using a regression with y_{it} on the LHS. Then the RHS will be the conditional probability that $y_{it} = 1$, plus an error term.

This is called a linear probability model:

$$y_{it} = \alpha_0 + \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + u_i + \varepsilon_{it} \qquad (1)$$

With panel data methods (*e.g.* within-group or random-effects), linear model implies:

$$E(y_{it} \mid \mathbf{z}_i, \mathbf{x}_{it}, u_i) \equiv \Pr(y_{it} = 1 \mid \mathbf{z}_i, \mathbf{x}_{it}, u_i) = P(\mathbf{z}_i, \mathbf{x}_{it}, u_i)$$

Disadvantages of the LPM

Model (1) requires:

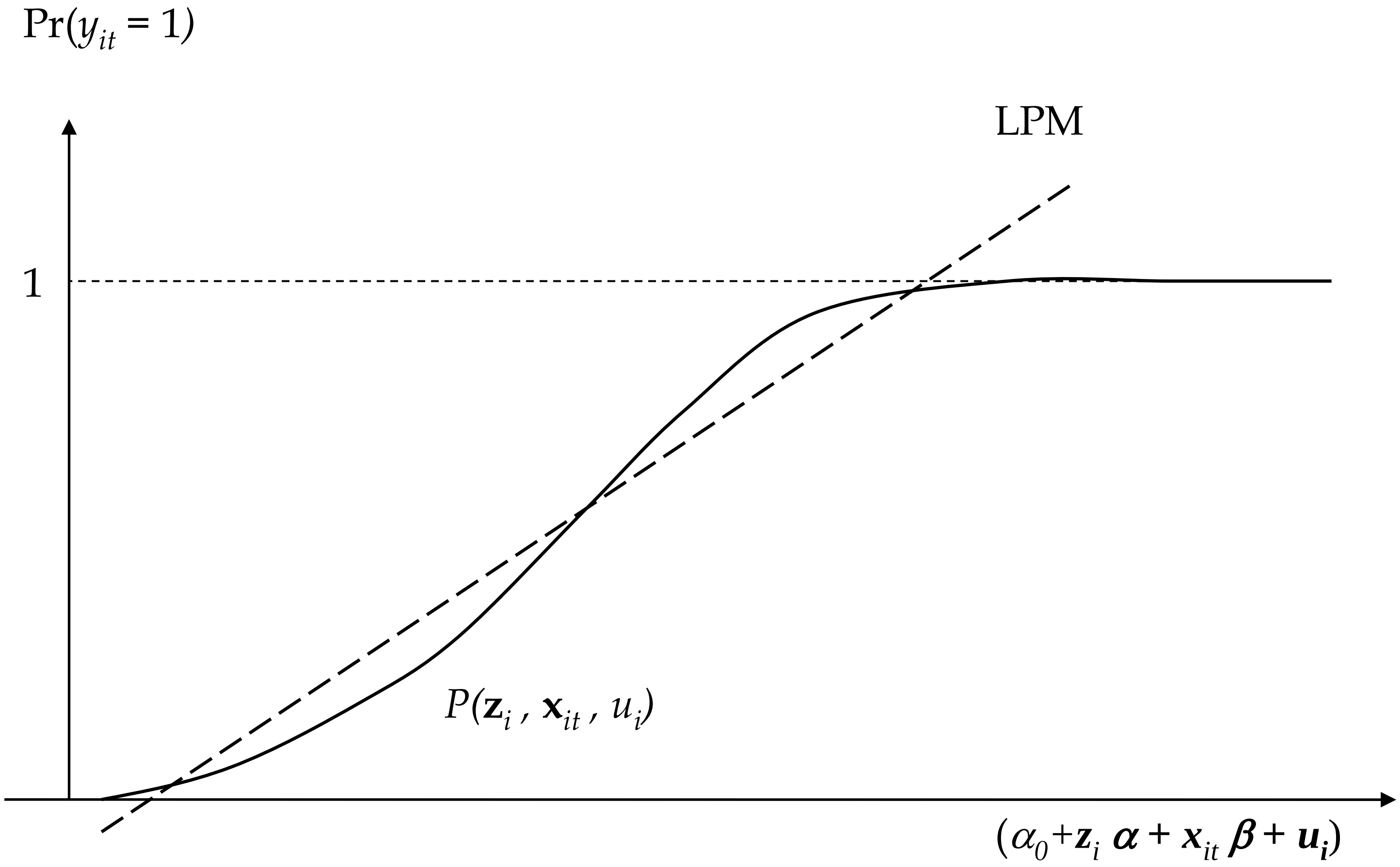
$$P(\mathbf{z}_i\, , \mathbf{x}_{it}\, , u_i) \;\approx\; \alpha_0 + \mathbf{z}_i\boldsymbol{\alpha} \;+\; \mathbf{x}_{it}\boldsymbol{\beta} \;+\; u_i$$

But this may fall outside the admissible [0, 1] interval.

Moreover, var(y_{it} | $\mathbf{z}_i$, $\mathbf{x}_{it}$, u_i) = $P(\mathbf{z}_i , \mathbf{x}_{it} , u_i)[1\text{-}P(\mathbf{z}_i , \mathbf{x}_{it} , u_i)]$ is not constant $\Rightarrow$ heteroskedasticity is a problem

[Despite its disadvantages, the panel LPM is simple to estimate and is often seen in applied work – but it's not an ideal choice.]

Why nonlinear models are needed



Latent regression models: the binary case (1)

To overcome the disadvantages of the LPM, use non-linear methods.

Define a latent (unobservable) continuous counterpart, y_{it}^*

Example from labour economics:

If y_{it}=1 defines employment, then:

y_{it}^* = best available wage – minimum acceptable wage.

Let y_{it}^* be generated by a linear regression structure:

$$y_{it}^* = \alpha_0 + \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + u_i + \varepsilon_{it}$$

Then employment is chosen whenever *available wage - acceptable wage* is positive:

$$y_{it} = 1 \quad \text{if and only if} \quad y_{it}^* > 0$$

Latent regression models: the binary case (2)

$$\Rightarrow \quad \Pr(y_{it} = 1 \mid \mathbf{z}_i, \mathbf{x}_{it}, u_i) = \Pr(\alpha_0 + \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + u_i + \varepsilon_{it} > 0)$$

$$= \Pr(-\varepsilon_{it} < [\alpha_0 + \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + u_i])$$

$$= F(\alpha_0 + \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + u_i)$$

where $F(.)$ is the distribution function of the random variable $-\varepsilon_{it}$

Probit model: assume ε_{it} has a normal distribution

$F(\,.\,) = \Phi(\,.\,) \Rightarrow$ df of the N(0,1) distribution

Logit (logistic regression) model: assume ε_{it} has a logistic distribution

$F(\varepsilon) = e^{\varepsilon}/[\,1+e^{\varepsilon}\,] \Rightarrow$ df of the logistic distribution

Example: RE probit model of PT work

```
xtprobit pt age female kid

Random-effects probit regression                Number of obs      =     59615
Group variable: pid                             Number of groups   =     10077
Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       5.9
                                                               max =        14
                                                Wald chi2(3)       =   3256.56
Log likelihood  = -15481.832                    Prob > chi2        =    0.0000
------------------------------------------------------------------------------
          pt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0234113   .0018799    12.45   0.000     .0197267    .0270959
      female |   2.936897   .0622869    47.15   0.000     2.814816    3.058977
         kid |   1.373893   .0340342    40.37   0.000     1.307187    1.440598
       _cons |  -5.018554   .0966839   -51.91   0.000    -5.208051   -4.829057
-------------+----------------------------------------------------------------
    /lnsig2u |   1.400743   .0367426                      1.328729    1.472757
-------------+----------------------------------------------------------------
     sigma_u |   2.014501    .037009                      1.943255    2.088359
         rho |   .8023018   .0058279                      .7906303    .8134761
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =  1.3e+04 Prob >= chibar2 = 0.000
```

Coefficients in non-linear models (1)

- Coefficients refer to the latent variable

  $$y_{it}^{*} = \alpha_0 + \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + u_i + \varepsilon_{it}$$

  not directly to the (probability of) outcome y_{it}.
- So coefficients are difficult to interpret directly.:
  - Can look at sign of coefficient (e.g. women are more likely to work PT than men)
  - Can make crude comparisons of coefficients if variables are in similar units (e.g. being female has a larger effect on PT work than having children, but don't know how much larger since depends on $\Phi(\,.\,)$).
  - Size of marginal effects depends on $F(\,.\,)$ transformation (Φ in probit).

Coefficients in non-linear models (2)

- Estimate of ρ (rho) gives proportion of latent variable error ($u_i + \varepsilon_{it}$) accounted for by individual effect u_i.

Understanding the results from binary latent regression models

In a linear regression model:

$$y_{it} = \alpha_0 + \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + u_i + \varepsilon_{it}$$

We can interpret the coefficients directly:

$\boldsymbol{\alpha}$ = (average) effect on y of increasing **z** by 1 unit

$\boldsymbol{\beta}$ = (average) effect on y of increasing **x** by 1 unit

These are known as the *marginal effects* of **z**, **x** on y

But , as seen, in nonlinear models, things are more complicated. In:

$$\Pr(y_{it} = 1) = F(\alpha_0 + \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + u_i)$$

$\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ aren't the effects on $\Pr(y_{it} = 1)$ of changing **z** or **x** by one unit $\Rightarrow$ so coefficients can't be directly interpreted

Understanding the results from binary latent regression models

There are different ways to summarise results:

- Predicting probabilities
- Calculating changes in probabilities (differences or ratios) due to discrete changes in explanatory variables
- Calculating ratios of odds corresponding to discrete changes in explanatory variables (and problems with this method…)

[Note can also calculate changes in probabilities based on infinitesimal (very small) changes in explanatory variables. This "derivative" method is traditionally presented in textbooks (standard formula) and used as the software default for continuous variables. "Discrete" versus "derivative" distinction can important in non-linear models. Derivative method may not be appropriate -- see additional slides.]

Some concepts for summarising results

Model: $\Pr(y_{it} = 1) = F(\alpha_0 + \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + u_i)$

(call this conditional probability P_{it})

Coefficients	=	α_0 , $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$
Predicted probability	=	P_{it}
Odds (O_{it})	=	$P_{it} / (1 - P_{it})$

For 2 people with different **z** and **x** –values, whose probabilities of y=1 are P_0 and P_1 :

Odds ratio	=	O_1 / O_0
Relative risk	=	P_1 / P_0

Relative risk and the odds ratio are often confused, but they are different

Marginal effects, relative risk and the odds ratio

Suppose person 0 has observable characteristics $\mathbf{z}_0$, $\mathbf{x}_0$ and unobservable characteristic u_0 ; then:

$$P_0 = F(\alpha_0 + \mathbf{z}_0\boldsymbol{\alpha} + \mathbf{x}_0\boldsymbol{\beta} + u_0)$$

Let's consider the effect of making a 1-unit change in (say) $\mathbf{z}$. This means inventing a new person with characteristics:

($\mathbf{z}_0$+1 , $\mathbf{x}_0$, u_0), for whom Pr(y=1) is:

$$P_1 = F(\alpha_0 + [\mathbf{z}_0+1]\boldsymbol{\alpha} + \mathbf{x}_0\boldsymbol{\beta} + u_0)$$

We can summarise the effect of this change in various ways:

- *Marginal effect* = $P_1 - P_0$
- *Relative risk* = P_1 / P_0
- *Odds ratio* = $[P_1 / (1 - P_1)] / [P_0 / (1 - P_0)]$
  $= [P_1 / P_0] \times [(1 - P_0) / (1 - P_1)]$

Other variables are "held constant" at their baseline values($\mathbf{x}_0$, u_0)

Example: RE probit

- Model of part-time work as a function of age, gender and presence of children.
- $\Pr(pt_{it} = 1) = F(\alpha_0 + \alpha_1 female_{it} + \beta_1 age_{it} + \beta_2 kid_{it} + u_i)$
  $= \Phi(\alpha_0 + \alpha_1 female_{it} + \beta_1 age_{it} + \beta_2 kid_{it} + u_i)$
- Use results to calculate $\Pr(pt_{it} = 1)$ for different people (setting $u_i = 0$ for simplicity).
- Use the `normal()` function (= Φ) in Stata. Manual method is flexible and good for illustrating concept.
- Also show `-margins-` command.
- What is the effect of presence of children on probability of working PT?

RE probit

```
. xtprobit pt age i.female i.kid
Random-effects probit regression                Number of obs      =     59615
Group variable: pid                             Number of groups   =     10077
Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       5.9
                                                               max =        14
                                                Wald chi2(3)       =   3256.56
Log likelihood  = -15481.832                    Prob > chi2        =    0.0000
------------------------------------------------------------------------------
          pt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0234113   .0018799    12.45   0.000     .0197267    .0270959
    1.female |   2.936897   .0622869    47.15   0.000     2.814816    3.058977
       1.kid |   1.373893   .0340342    40.37   0.000     1.307187    1.440598
       _cons |  -5.018554   .0966839   -51.91   0.000    -5.208051   -4.829057
-------------+----------------------------------------------------------------
    /lnsig2u |   1.400743   .0367426                      1.328729    1.472757
-------------+----------------------------------------------------------------
     sigma_u |   2.014501    .037009                      1.943255    2.088359
         rho |   .8023018   .0058279                      .7906303    .8134761
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =  1.3e+04 Prob >= chibar2 = 0.000
```

Predicted probabilities

- Calculate predicted probability that a 40 year old woman with no children works PT (P_0).

```
nlcom normal(_b[_cons] + _b[age]*40 + _b[1.female])
------------------------------------------------------------------------------
          pt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |   .1260621    .007955    15.85   0.000     .1104705    .1416537
------------------------------------------------------------------------------
```

- Calculate predicted probability for similar woman with children (P_1).

```
nlcom normal(_b[_cons] + _b[age]*40 + _b[1.female] + _b[1.kid])
------------------------------------------------------------------------------
          pt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |    .590444   .0159042    37.13   0.000     .5592724    .6216156
------------------------------------------------------------------------------ -
```

Marginal effects (1)

- Marginal effect is $P_1 - P_0 = 0.590 - 0.126 = 0.464$. Can also calculate directly (gives standard error of ME):

```
nlcom normal(_b[_cons] + _b[age]*40 + _b[1.female] + _b[1.kid]) -
  normal(_b[_cons] + _b[age]*40 + _b[1.female])

------------------------------------------------------------------------------
          pt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |   .4643819   .0124861    37.19   0.000     .4399096    .4888543
------------------------------------------------------------------------------
```

Marginal effects (2)

- Can also calculate ME using `-margins-` (new in version 11, replaces `-mfx-`), though it may be less flexible (see later):

```
. margins, dydx(kid) predict(pu0) at(age=40 female=1 kid=0)

Conditional marginal effects                      Number of obs   =      59615
Model VCE    : OIM

Expression   : Pr(pt=1 assuming u_i=0), predict(pu0)
dy/dx w.r.t. : 1.kid
at           : age             =          40
               female          =           1
               kid             =           0

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.kid |   .4643819   .0124861    37.19   0.000     .4399096    .4888543
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

Relative risk

- Relative risk is P_1 / P_0 = 0.590 / 0.126 = 4.683. Calculating directly (note difficult to do using `-margins-`):

```
nlcom normal(_b[_cons] + _b[age]*40 + _b[1.female] + _b[1.kid]) /
  normal(_b[_cons] + _b[age]*40 + _b[1.female])

------------------------------------------------------------------------------
          pt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |   4.683756   .2366925    19.79   0.000     4.219847    5.147665
------------------------------------------------------------------------------
```

Logistic regression and the odds ratio

In the logit model:

$$P_0 = exp(\alpha_0 + \mathbf{z}_0 \boldsymbol{\alpha} + \mathbf{x}_0 \boldsymbol{\beta} + u_0) \ / \ [\,1 + exp(\alpha_0 + \mathbf{z}_0 \boldsymbol{\alpha} + \mathbf{x}_0 \boldsymbol{\beta} + u_0)\,]$$

$$P_1 = exp(\alpha_0 + [\mathbf{z}_0 + 1]\boldsymbol{\alpha} + \mathbf{x}_0 \boldsymbol{\beta} + u_0) \ / \ [\,1 + exp(\alpha_0 + [\mathbf{z}_0 + 1]\boldsymbol{\alpha} + \mathbf{x}_0 \boldsymbol{\beta} + u_0)\,]$$

$$Odds\ ratio = [\,P_1 \ / \ (1 - P_1)\,] \ / \ [\,P_0 \ / \ (1 - P_0)\,]$$

$$= [\,exp(\alpha_0 + [\mathbf{z}_0 + 1]\boldsymbol{\alpha} + \mathbf{x}_0 \boldsymbol{\beta} + u_0)\,] \ / \ [\,exp(\alpha_0 + \mathbf{z}_0 \boldsymbol{\alpha} + \mathbf{x}_0 \boldsymbol{\beta} + u_0)\,]$$

$$= [\,exp(\alpha_0 + \mathbf{z}_0 \boldsymbol{\alpha} + \mathbf{x}_0 \boldsymbol{\beta} + u_0) \times exp(1 \times \boldsymbol{\alpha})\,] \ / \ [\,exp(\alpha_0 + \mathbf{z}_0 \boldsymbol{\alpha} + \mathbf{x}_0 \boldsymbol{\beta} + u_0)\,]$$

$$= exp(\boldsymbol{\alpha})$$

The odds ratio is usually only quoted in relation to logit results. It is hard to interpret and very often gets misinterpreted. It gives the proportionate effect of a 1-unit change in a variable on the ***odds*** , not the probability Pr(y=1).

Misinterpretation of odds ratios

Check that you understand the error in the following quotation:

"The odds ratio of 1.3689 for females […] indicates that, controlling for the effects of the other explanatory variables, females are 37% more likely to be in poverty than males. Stated differently, the probability of being in poverty is 1.37 times greater for females than for males."

(W. H. Crown, *Statistical Models for the Social and Behavioural Sciences: Multiple Regression and Limited Dependent Variable Models*. London: Praeger, 1998)

It isn't possible to calculate the relative risk or the marginal effect on the response probability, from knowledge of the odds ratio alone.

What would be the relative risk and marginal effect if the predicted probability for males is 0.2? What if it's 0.001? What if it's 0.8?

Choice of reference/baseline values (1)

- The choice of baseline values ($\mathbf{z}_0$, $\mathbf{x}_0$ and u_0) is important. Can have big impact on calculated effects.
- Needs to be guided by research question. *Who* are we interested in calculating probabilities or marginal effects for?
- Possibilities:
  - Mean $\mathbf{x}$, $\mathbf{z}$ (and mean u[=0]). Often seen in lit, and used as the default by several software packages (was default for `mfx` but not for `margins`). Easy to do with `margins`. But this represents a synthetic, hybrid person that doesn't exist…

Choice of reference/baseline values (2)

- Possibilities contd:
  - Actual **x** and **z** for each person in sample (or a sub-sample), then take average to give mean effect. Easy to do with `margins`.
  - Values of $\mathbf{z}_0$, $\mathbf{x}_0$ chosen to represent different types of person. Can do with `margins` or `nlcom`.
  - Set u_0=0 (mean of u_i) or use representative values, e.g. +/- 1 standard deviation (σ_u), to see how probabilities and marginal effects depend on unobserved heterogeneity . Need to use `nlcom` since `margins` restricts $u_0 = 0$.
  - It is also possible to calculate *average partial effects* (APE) which allow for the average effect of u_i (see additional slides).

Options for presenting results from binary response models (1)

- Predict and compare probabilities for different combinations of **x** and **z** (and *u*), representing different types of person or counterfactual scenarios. Use `nlcom` or `margins`.
- Calculate marginal effects for changes in **x** or **z** (and *u*). Use `nlcom` or `margins`.
  - `nlcom` is usually more flexible. Can calculate marginal effects (or relative risks) for any combination of changes in **x** or **z**. Note that `margins` only allows one characteristic to change at a time.
  - `margins` uses the "derivative method " to calculate marginal effects for non-categorical variables, i.e. calculates effects of infinitesimal change, not discrete change. May be inappropriate in non-linear models – see additional slides.

Options for presenting results from binary response models (2)

- All these methods are difficult with the FE logit, as we don't estimate the (distribution of) individual effects or even a constant. But:
  - Can still obtain odds ratios
  - Can get a crude measure of marginal effects by evaluating at the sample proportion of 1s. Then $\partial P_{it} / \partial x_{jit} = \bar{y}(1-\bar{y})\beta_j$

Fixed effects models – some issues

- To deal with individual effects in linear FE models, we can:
  - estimate individual effects u_i (LSDV).
  - difference out individual effects u_i (first differences or within transform) .

  Estimates of **β** are are unbiased in both cases.
- But in non-linear FE models:
  - There's in general no short-cut method that avoids also estimating u_i $\Rightarrow$ the "incidental parameters problem. And estimated coefficients are "biased" in this case.
  - Can't remove the individual effects u_i by simple differencing as in within-group regression

Conditional ML estimation

- CML (as applied here) is a way of condensing the likelihood function into a form which does not depend on u_i but does depend on $\boldsymbol{\beta}$**.**
- Then CML is consistent (loosely speaking, unbiased in a large sample) for $\boldsymbol{\beta}$.
- But CML is very model specific as it is based on a technical "trick" that is only applicable in a few cases, *e.g.*:
    - logit models
    - Poisson model

Fixed effects (or conditional) logit

Model: $\Pr(y_{it} = 1) = F(\alpha_0 + \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + u_i)$,
where $F(\,.\,)$ is the logistic form

Roughly speaking, the method works as follows:

- Work with the subsample of individuals for whom there is some change in y_{it} during the observation period $\Rightarrow$ so we sacrifice information on any individuals displaying no change in y
- The changes in the covariates $\mathbf{x}_{it}$ (*i.e.* variable differences like $\mathbf{x}_{it} - \mathbf{x}_{it-1}$) are then used in a modified logit analysis to explain the changes in the observed sequence of outcomes $y_{i1} \ldots y_{iT}$.
- Note that differencing the covariates removes any variables that are constant over time (*e.g.* gender, birth year, etc.), so $\boldsymbol{\alpha}$ can't be estimated
- But it also removes u_i , so we don't have to assume anything about u_i $\Rightarrow$ so FE logit is more robust than RE logit

Conditional logit

Subsume $\mathbf{z}_i$ in $\mathbf{x}_{it}$ for notational simplicity.

If we try to estimate the u_i using individual-specific dummy variables, there is no simplification analogous to within-group regression.

Moreover, the number of parameters $\rightarrow \infty$ with n, so the MLDV estimator is not consistent.

Log-likelihood for the logit model for individual i conditional on u_i:

$$L(\beta, u_1 \ldots u_n) = \sum_{t=1}^{T_i} (1 - y_{it}) \ln\left(\frac{1}{1 + e^{\mathbf{x}_{it}\boldsymbol{\beta} + u_i}}\right) + \sum_{t=1}^{T_i} y_{it} \ln\left(\frac{e^{\mathbf{x}_{it}\boldsymbol{\beta} + u_i}}{1 + e^{\mathbf{x}_{it}\boldsymbol{\beta} + u_i}}\right)$$

The statistic $\sum_t y_{it}$ is a sufficient statistic for u_i: $\Pr(\mathbf{y}_i \mid \sum_t y_{it})$ does not depend on u_i.

Conditional logit (continued)

Example $T_i = 2$; $\sum_t y_{it}$ can take values 0, 1, 2.

1. Conditional on $\sum_t y_{it} = 0$:

$$\Pr(y_{i1} = 0, y_{i2} = 0 \mid y_{i1} + y_{i2} = 0) = 1$$
$$\Pr(y_{i1} = 0, y_{i2} = 1 \mid y_{i1} + y_{i2} = 0) = 0$$
$$\Pr(y_{i1} = 1, y_{i2} = 0 \mid y_{i1} + y_{i2} = 0) = 0$$
$$\Pr(y_{i1} = 1, y_{i2} = 1 \mid y_{i1} + y_{i2} = 0) = 0$$

Does not depend on $\boldsymbol{\beta}$ so can drop from likelihood function.

2. Conditional on $\sum_t y_{it} = 2$:

$$\Pr(y_{i1} = 1, y_{i2} = 1 \mid y_{i1} + y_{i2} = 2) = 1$$

Etc

Again, does not depend on $\boldsymbol{\beta}$ so can drop from likelihood function.

3. Only cases with $\sum_t y_{it} = 1$ are of interest, specifically:

$\Pr(y_{i1} = 1, y_{i2} = 0 \mid y_{i1} + y_{i2} = 1)$ and $\Pr(y_{i1} = 0, y_{i2} = 1 \mid y_{i1} + y_{i2} = 1)$.

Work through algebra for $\Pr(y_{i1} = 1, y_{i2} = 0 \mid y_{i1} + y_{i2} = 1)$

Conditional logit (continued)

Probability of the conditioning event:

$$\Pr(\textstyle\sum_t y_{it} = 1) = \Pr(y_{i1} = 1, y_{i2} = 0) + \Pr(y_{i1} = 0, y_{i2} = 1)$$

$$= P_{i1}(1-P_{i2}) + (1-P_{i1})P_{i2}$$

$$= \frac{e^{\mathbf{x}_{i1}\boldsymbol{\beta}+u_i} + e^{\mathbf{x}_{i2}\boldsymbol{\beta}+u_i}}{(1+e^{\mathbf{x}_{i1}\boldsymbol{\beta}+u_i})(1+e^{\mathbf{x}_{i2}\boldsymbol{\beta}+u_i})}$$

Conditional probability:

$$\Pr(y_{i1} = 1, y_{i2} = 0 \mid y_{i1} + y_{i2} = 1) = \frac{\Pr(y_{i1} = 1, y_{i2} = 0)}{\Pr(y_{i1} + y_{i2} = 1)}$$

$$= \frac{e^{\mathbf{x}_{i1}\boldsymbol{\beta}+u_i}}{e^{\mathbf{x}_{i1}\boldsymbol{\beta}+u_i} + e^{\mathbf{x}_{i2}\boldsymbol{\beta}+u_i}} = \frac{e^{\mathbf{x}_{i1}\boldsymbol{\beta}}}{e^{\mathbf{x}_{i1}\boldsymbol{\beta}} + e^{\mathbf{x}_{i2}\boldsymbol{\beta}}} = \frac{e^{(\mathbf{x}_{i1}-\mathbf{x}_{i2})\boldsymbol{\beta}}}{1+e^{(\mathbf{x}_{i1}-\mathbf{x}_{i2})\boldsymbol{\beta}}}$$

$\Rightarrow u_i$ is eliminated by conditioning on $\sum_t y_{it}$

By similar working/symmetry $\Pr(y_{i1} = 0, y_{i2} = 1 \mid y_{i1} + y_{i2} = 1) = \dfrac{e^{-(\mathbf{x}_{i1}-\mathbf{x}_{i2})\boldsymbol{\beta}}}{1+e^{-(\mathbf{x}_{i1}-\mathbf{x}_{i2})\boldsymbol{\beta}}}$

Conditional logit (continued)

With $T = 2$, the conditional log-likelihood is:

$$L(\boldsymbol{\beta}) = \sum_{i:\Sigma y=1} d_i(\mathbf{x}_{i1} - \mathbf{x}_{i2})\boldsymbol{\beta} - \ln(1 + e^{(\mathbf{x}_{i1} - \mathbf{x}_{i2})\boldsymbol{\beta}})$$

where $d_i = 1$ if $y_{i1} = 1$, $y_{i2} = 0$ and 0 if $y_{i1} = 0$, $y_{i2} = 1$.

Note that, if $\mathbf{x}_{it}$ contains time-invariant covariates (*i.e.* $\mathbf{z}_i$), these disappear from $(\mathbf{x}_{i1}-\mathbf{x}_{i2}) \Rightarrow \boldsymbol{\alpha}$ cannot be estimated.

In general, conditional logit only uses data from individuals who experience change in y_{it} over time. This sacrifices sample variation.

- The same conditioning approach does not work with probit and other functional forms, nor with general dynamic models
- But it can be generalised to:
    - unordered multinomial logit models
    - ordered logit models with more than two outcomes.

Random effects logit/probit

If we want to:

- estimate the coefficients of $\mathbf{z}_i$
- use a non-logistic form
- allow for dynamic adjustment *(i.e.* use the lagged value $y_{it\text{-}1}$ *as* an explanatory variable)

then conditional likelihood is not available. The random effects approach is a natural solution.

Random effects logit/probit

Consider the basic model:

$$y_{it}^{*} = \alpha_0 + \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + u_i + \varepsilon_{it}$$

$$y_{it} = 1 \quad \text{if and only if } y_{it}^{*} > 0$$

Make standard random effects assumptions (including independence of ($\mathbf{z}_i$, $\mathbf{x}_{it}$) and u_i).

Define probabilities $P_{it}(u_i) \equiv \Pr(y_{it} \,|\, \mathbf{z}_i, \mathbf{x}_{it}, u_i)$ of the form:

$F(\alpha_0 + \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + u_i)$ for $y_{it} = 1$

$1 - F(\alpha_0 + \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + u_i)$ for $y_{it} = 0$.

Since the ε_{it} are independent, the joint probability of observing (y_{i1}, y_{i1},…, y_{iTi}) conditional on u_i (and $\mathbf{z}_i$, $\mathbf{x}_{it}$) is the product of these conditional probabilities.

Random effects logit/probit

Likelihood function for individual i, conditional on u_i (and $\mathbf{z}_i$, $\mathbf{x}_{it}$) is:

$$L_i(u_i) = \prod_{t=1}^{T_i} P_{it}(u_i)$$

Marginalise with respect to (or "average out") u_i:

$$L_i = E\left(\prod_{t=1}^{T_i} P_{it}(u_i)\right) = \int_{-\infty}^{\infty} \prod_{t=1}^{T_i} P_{it}(u)g(u)du \qquad (1)$$

Where $g(u)$ is an assumed density for u, e.g. normal (Gaussian): $g(u) = \sigma_u^{-1}\phi(u/\sigma_u)$.

Evaluation of the likelihood function requires the integral in (1) to be approximated numerically by a quadrature algorithm.

This is implemented in Stata (`xtprobit`, `xtlogit`), but computing run times are quite long.

Is the zero-correlation assumption valid? The Hausman test

- A Hausman test can be used to compare conditional logit estimates with the random-effects logit which assumes independence between u_i and $(\mathbf{z}_i, \mathbf{X}_i)$.
- Null hypothesis is H_0: u_i and $(\mathbf{z}_i, \mathbf{X}_i)$ are independent.
- Alternative hypothesis is H_1: u_i and $(\mathbf{z}_i, \mathbf{X}_i)$ are not independent (implies we should use CL).
- $\hat{\boldsymbol{\beta}}_{CL}$ is consistent under H_0 and H_1, but inefficient under H_0 (since only uses information on changers).
- $\hat{\boldsymbol{\beta}}_{RE}$ is consistent and efficient under H_0, but inconsistent under H_1.
- Test statistic:

$$S = (\hat{\boldsymbol{\beta}}_{CL} - \hat{\boldsymbol{\beta}}_{RE})'(\mathrm{var}(\hat{\boldsymbol{\beta}}_{CL}) - \mathrm{var}(\hat{\boldsymbol{\beta}}_{RE}))(\hat{\boldsymbol{\beta}}_{CL} - \hat{\boldsymbol{\beta}}_{RE})$$

(distributed as χ^2 if H_0 is correct, with df equal to the no. of coefficients in $\boldsymbol{\beta}$)

Example: FE (conditional) logit

```
. xtlogit pt age female kid, fe
note: multiple positive outcomes within groups encountered.
note: 8060 groups (43047 obs) dropped because of all positive or
      all negative outcomes.
note: female omitted because of no within-group variance.
Conditional fixed-effects logistic regression   Number of obs      =     16568
Group variable: pid                             Number of groups   =      2017
                                                Obs per group: min =         2
                                                               avg =       8.2
                                                               max =        14
                                                LR chi2(2)         =   1492.98
Log likelihood  = -5725.2774                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
          pt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0024084   .0056486     0.43   0.670    -.0086627    .0134796
      female |  (omitted)
         kid |   2.391404   .0706994    33.82   0.000     2.252836    2.529973
------------------------------------------------------------------------------
```

Example: RE logit

```
. xtlogit pt age female kid, re
Random-effects logistic regression              Number of obs      =     59615
Group variable: pid                             Number of groups   =     10077
Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       5.9
                                                               max =        14
                                                Wald chi2(3)       =   2818.31
Log likelihood  = -15440.418                    Prob > chi2        =    0.0000
------------------------------------------------------------------------------
          pt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0436602    .003382    12.91   0.000     .0370315    .0502889
      female |   5.251262   .1186133    44.27   0.000     5.018784     5.48374
         kid |    2.53694   .0641967    39.52   0.000     2.411117    2.662763
       _cons |  -9.064061   .1805762   -50.20   0.000    -9.417984   -8.710138
-------------+----------------------------------------------------------------
    /lnsig2u |   2.539976   .0396153                      2.462331     2.61762
-------------+----------------------------------------------------------------
     sigma_u |   3.560809   .0705312                       3.42522    3.701766
         rho |   .7939871   .0064799                      .7809966     .806398
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =  1.3e+04 Prob >= chibar2 = 0.000
```

Example: Hausman test comparing FE and RE logit

```
. hausman fixed random

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed        random       Difference          S.E.
-------------+----------------------------------------------------------------
         age |    .0024084     .0436602       -.0412518        .0045243
         kid |    2.391404      2.53694       -.1455357        .0296175
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtlogit
            B = inconsistent under Ha, efficient under Ho; obtained from xtlogit

    Test:  Ho:  difference in coefficients not systematic

                  chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =      165.80
                Prob>chi2 =      0.0000
```

Additional slides

Quick intro to maximum likelihood (1)

- Non-linear models are usually estimated by maximum likelihood (ML).
- ML works as follows:
  1. Write down likelihood (probability) of observing outcomes in the data (according to the model!)
  2. Likelihood is a function of the unknown parameters (coefficients) of the model.
  3. Use computer algorithm to choose the parameter values that give the highest likelihood. These are our estimates.
- Example of cross-sectional probit: we know that $\Pr(y_i{=}1) = \Phi(\mathbf{x}_i\,\boldsymbol{\beta})$ and $\Pr(y_i{=}0) = 1\text{-}\Phi(\mathbf{x}_i\,\boldsymbol{\beta})$ [note $\mathbf{x}_i$ includes constant].

Quick intro to maximum likelihood (2)

- Say outcomes are 1, 1, 0, …, 1. Probability of observing these data (assuming independent observations) is:

  $\Pr(y_1=1) \times \Pr(y_2=1) \times \Pr(y_3=0), \ldots, \times \Pr(y_N=1)$

  $= \Phi(\mathbf{x}_1\boldsymbol{\beta})\ \Phi(\mathbf{x}_2\boldsymbol{\beta})\ [1-\Phi(\mathbf{x}_3\boldsymbol{\beta})], \ldots, \Phi(\mathbf{x}_N\boldsymbol{\beta})$

- The (cross-sectional) probit likelihood function can be written:

$$L(\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{x}) = \prod_{i=1}^{N} \Phi(\mathbf{x}_i\boldsymbol{\beta})^{y_i} [1 - \Phi(\mathbf{x}_i\boldsymbol{\beta})]^{1-y_i}$$

Quick intro to maximum likelihood (3)

- For convenience, we normally takes logs so the log likelihood is the sum of individual likelihoods:

$$\log L(\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{x}) = \sum_{i=1}^{N} y_i \ln \Phi(\mathbf{x}_i \boldsymbol{\beta}) + (1 - y_i) \ln[1 - \Phi(\mathbf{x}_i \boldsymbol{\beta})]$$

- Stata uses interative algorithm to choose $\boldsymbol{\beta}$ to maximise log L.
- Note random-effects version of probit log L is more complicated, as we see later…

Predicting probabilities

- Can use `predict` in Stata to predict probability of positive outcome for each observation in sample (mainly useful for evaluating range of probabilities predicted by model).
- To predict probability for a representative person, plug appropriate values into the $F(.)$ formula.
- In the probit model:

  $\Pr(y = 1) = \Phi(\alpha_0 + \mathbf{z}_0\boldsymbol{\alpha} + \mathbf{x}_0\boldsymbol{\beta} + u_0)$, where $\mathbf{z}_0$ etc are the representative values chosen (how would you choose u_0?).

  Use the Stata `normal()` function (= Φ) to evaluate.

Predicting probabilities contd.

- In the logit model:

  $\Pr(y=1) = \exp(\alpha_0+\mathbf{z}_0\boldsymbol{\alpha}+\mathbf{x}_0\boldsymbol{\beta}+u_0)/[1+\exp(\alpha_0+\mathbf{z}_0\boldsymbol{\alpha}+\mathbf{x}_0\boldsymbol{\beta}+u_0)]$

  Use the Stata `exp()` function to evaluate.
- Example: probit model of part-time work as a function of age, gender and presence of children.
- $\Pr(pt_{it}=1) = F(\alpha_0 + \alpha_1 female_{it} + \beta_1 age_{it} + \beta_2 kid_{it} + u_i)$

  $= \Phi(\alpha_0 + \alpha_1 female_{it} + \beta_1 age_{it} + \beta_2 kid_{it} + u_i)$

Predicted probabilities

- After estimation, calculate predicted probability of working PT for a 40 year old woman with no children (setting $u_i = 0$ for simplicity).
- Use `nlcom`, which automatically generates standard errors etc.
- Access coefficients using `_b[variable name]` (note `_b` refers to estimate of β).

```
. nlcom normal(_b[_cons] + _b[age]*40 + _b[female])
------------------------------------------------------------------------------
          pt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |   .1260621    .007955    15.85   0.000     .1104705    .1416537
------------------------------------------------------------------------------
```

Marginal effects based on small changes in explanatory variables

- We have focused on marginal effects as the change in $\Pr(y = 1)$ corresponding to a given (discrete) change in **x** or **z**.
- However, marginal effects are often reported as the change in $\Pr(y = 1)$ corresponding to a very small (infinitesimal) change in **x** or **z**, scaled up to represent a 1 unit change.
- This "derivative" method has traditionally been a popular way to present results (common in textbooks), partly because the effects can be calculated directly using a standard formula (see following slides). Stata command -`margins`- will do this for us.

Marginal effects based on small changes in explanatory variables

- Scaling up the effect due to an infinitesimal change is fine in linear models because we fit a straight line through the data, e.g. we can say that $\partial y/\partial x = 0.3$ means that a one unit change in x would change y by 0.3
- But scaling up like this is not generally correct in non-linear models if the change you wish to consider is not small, e.g. change in dummy variable (0 to 1) or increase of discrete variable (going from 2 to 3 children may not be a small change!). The problem arises because we fit a curved function to the data (normal or logistic CDF).

Marginal effects based on small changes in explanatory variables

- No hard and fast rules about difference between the 2 methods (will also depend on size of coefficients and baseline probability). But it is safest to use the discrete method (difference in probability before and after change).
- `-margins-` recognises dummy variables (if specified using factor notation, i.varname) and calculates effect due to 0 to 1 change. But `–margins-` calculates marginal effect based on infinitesimal change for all other variables (including discrete variables with >2 categories).

Example: PT work and number of children

```
xtprobit pt age i.female nchild
Random-effects probit regression                Number of obs      =     59615
Group variable: pid                             Number of groups   =     10077
Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       5.9
                                                               max =        14
                                                Wald chi2(3)       =   3218.70
Log likelihood  = -15452.175                    Prob > chi2        =    0.0000
------------------------------------------------------------------------------
          pt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0253192   .0019068    13.28   0.000      .021582    .0290565
    1.female |   3.008655    .063591    47.31   0.000     2.884019    3.133291
      nchild |   .7476797   .0183303    40.79   0.000     .7117529    .7836064
       _cons |  -5.117774   .0996596   -51.35   0.000    -5.313103   -4.922445
-------------+----------------------------------------------------------------
    /lnsig2u |    1.42899   .0367436                      1.356974    1.501007
-------------+----------------------------------------------------------------
     sigma_u |   2.043155   .0375365                      1.970894    2.118066
         rho |    .806744   .0057286                      .7952675    .8177246
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =  1.4e+04 Prob >= chibar2 = 0.000
```

Example: PT work and number of children

Effect of going from 2 to 3 children

```
nlcom normal(_b[_cons] + _b[age]*40 + _b[1.female] + _b[nchild]*3) ///
> - normal(_b[_cons] + _b[age]*40 + _b[1.female]+ _b[nchild]*2)
------------------------------------------------------------------------------
          pt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |    .219188    .005857    37.42   0.000     .2077086    .2306675
------------------------------------------------------------------------------
```

Effect based on infinitesimal increase from 2 children (derivative method)

```
margins, dydx(nchild) predict(pu0) at(age=40 female=1 nchild=2)
Conditional marginal effects                      Number of obs   =      59615
Expression   : Pr(pt=1 assuming u_i=0), predict(pu0)
dy/dx w.r.t. : nchild
------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      nchild |   .2754571   .0057256    48.11   0.000     .2642351    .2866791
------------------------------------------------------------------------------
```

Marginal effects of continuous variables (derivative method)

- In the LPM, the marginal effect of an increase in a variable on the conditional probability that $y_{it} = 1$ is just its coefficient. Formally $\partial P(\mathbf{x}_{it}, u_i) / \partial \mathrm{x}_{jit} = \beta_j$ (where $\mathbf{z}_i$ is absorbed into $\mathbf{x}_{it}$ for brevity)
- Note the marginal effect in the LPM does not depend on the values of other covariates, or the individual effect. So the ME is the same for everyone.
- This is not generally true in non-linear models:

$$\partial P(\mathbf{x}_{it}, u_i) / \partial \mathrm{x}_{jit} = \partial F(\alpha_0 + \mathbf{x}_{it}\boldsymbol{\beta} + u_i) / \partial \mathrm{x}_{jit}$$

$$= f(\alpha_0 + \mathbf{x}_{it}\boldsymbol{\beta} + u_i)\, \beta_j$$

Marginal effects of continuous variables (derivative method)

- For the probit model:

$$\partial P(\mathbf{x}_{it}, u_i) / \partial \mathrm{x}_{jit} = \partial \Phi(\alpha_0 + \mathbf{x}_{it}\boldsymbol{\beta} + u_i) / \partial \mathrm{x}_{jit}$$

$$= \beta_j \, \phi\, (\alpha_0 + \mathbf{x}_{it}\boldsymbol{\beta} + u_i)$$

- For the logit model:

$$\partial P(\mathbf{x}_{it}, u_i) / \partial \mathrm{x}_{jit}$$

$$= \partial\, [\exp(\alpha_0 + \mathbf{x}_{it}\boldsymbol{\beta} + u_i) / (1 + \exp(\alpha_0 + \mathbf{x}_{it}\boldsymbol{\beta} + u_i))] / \partial \mathrm{x}_{jit}$$

$$= \beta_j \exp(\alpha_0 + \mathbf{x}_{it}\boldsymbol{\beta} + u_i) / (1 + \exp(\alpha_0 + \mathbf{x}_{it}\boldsymbol{\beta} + u_i))^2$$

$$= \beta_j \exp(\alpha_0 + \mathbf{x}_{it}\boldsymbol{\beta} + u_i) / (1 + \exp(\alpha_0 + \mathbf{x}_{it}\boldsymbol{\beta} + u_i)$$

$$*[1 - \exp(\alpha_0 + \mathbf{x}_{it}\boldsymbol{\beta} + u_i) / (1 + \exp(\alpha_0 + \mathbf{x}_{it}\boldsymbol{\beta} + u_i)]$$

$$= \beta_j P(\mathbf{x}_{it}, u_i)(1 - P(\mathbf{x}_{it}, u_i))$$

Marginal effects of continuous variables (derivative method)

- Marginal effect is coefficient multiplied by the density function (normal for probit, logistic for logit), evaluated at the base values of **x**.
- So marginal effects depend on covariates and individual effects. And usually we don't estimate the individual effects directly!
- Note we can still compare the relative effects of variables (since $f(.)$ cancels out). So the ratio of MEs due to x_j and x_k is β_j / β_k . Doesn't depend on value of latent variable.

Odds ratios in the logit model

- The logit model can also be expressed as a function of odds ($\equiv P/(1 - P)$).

  $P(\mathbf{x}_{it}, u_i) / (1-P(\mathbf{x}_{it}, u_i)) = \exp(\alpha_0+\mathbf{x}_{it}\boldsymbol{\beta}+u_i)$
- If x_{jit} increases by 1 unit, holding all else constant, the ratio of the new to the old odds is $\exp(\beta_j)$.
- Equivalently, β_j is the marginal effect on the log odds.
- Both are independent of $\mathbf{x}$ and u_i but how easy to interpret?

Average partial effects (1)

- Consider the RE probit.
- We are potentially interested in:
  - probability given by the model. This is:
    $\Pr(y_{it} \mid \mathbf{z}_i, \mathbf{x}_{it}, u_i) = \Phi(\alpha_0 + \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + u_i)$
  - marginal effect of increasing x_{jit}. This is:
    $\partial \Pr(y_{it} \mid \mathbf{z}_i, \mathbf{x}_{it}, u_i) / \partial x_{jit} = \beta_j \phi(\alpha_0 + \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + u_i)$.

Average partial effects (2)

- On average, $u_i = 0$ in the population (by construction). However, setting $u_i = 0$ will generate the wrong probabilities/MEs for given individuals, and also probably the wrong average probabilities / MEs. [See "warning" in the Stata manual!]
- Average partial effects (APE) are the marginal effects evaluated at a given **z** and **x**, but allowing for the effect of u_i averaged over its whole distribution.

Average partial effects (3)

- Can show that APE = $(\beta_j/\sigma)\phi([\alpha_0+\mathbf{z}_i\boldsymbol{\alpha}+\mathbf{x}_{it}\boldsymbol{\beta}]/\sigma)$. The scaling factor $\sigma = \sqrt{(1+\sigma_u^2)}$ is the standard deviation of the composite error in the probit.
- Average probability = $\Phi([\alpha_0+\mathbf{z}_i\boldsymbol{\alpha}+\mathbf{x}_{it}\boldsymbol{\beta}]/\sigma)$.
- These can be evaluated in Stata using the `normal()` (=Φ) and `normalden()` (= ϕ) functions, for any desired value of **z** and **x**.