DISCRETE RANDOM VARIABLES

Discreteness

- ***Inherent discreteness*** might involve transitions between states (*e.g.* dividend/no dividend, investment / no investment)
- Sometimes there are no two-way transitions (e.g. default)
- ***Observational discreteness*** is an artefact of the observation process (*e.g.* CAP bucket FDIC)

[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.* value of investment 0 for no-trigger firms, any positive value for trigger firms)

- ***Truncation*** involves discarding part of the population

(*e.g.* SME targeted samples, Listed firms)

- ***Count variables*** are the outcome of some counting process

(*e.g.* the number of capital types firm invests in, or the number of product types a country exports)

- ***Binary variables*** reflect a distinction between two states

(*e.g.* default / no default, export / no export, dividend / no dividend)

- ***Ordinal variables*** are ordered variables, typically taking more than two values

(*e.g.* CAP 1-4, SME size 1-4)

- ***Unordered variables*** reflect outcomes which are discrete but with no natural ordering (*e.g.* bank specialization)

Binary response models

Binary models

Dependent variable is

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

This describes:

- situations of choice between 2 alternatives
- Binary outcomes are outcomes with two possible values, commonly referred to as *success* and *failure*.
- The outcome of interest (success) is commonly scored "1" if it occurs, otherwise "0" (failure).

E.g. suppose:

- $\mathbf{y}_i = (0, 0, 0, 0, 1, 1, 1, 0, 1, 1)$ is an annual panel observation
- 0 indicates no dividend was paid, 1 indicates dividend was paid

Then $\mathbf{y}_i$ represents a history of 4 years' of zero dividend pay out followed by 3 years' positive dividend, followed by 1 year's no dividend then 2 years' positive dividend.

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:

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

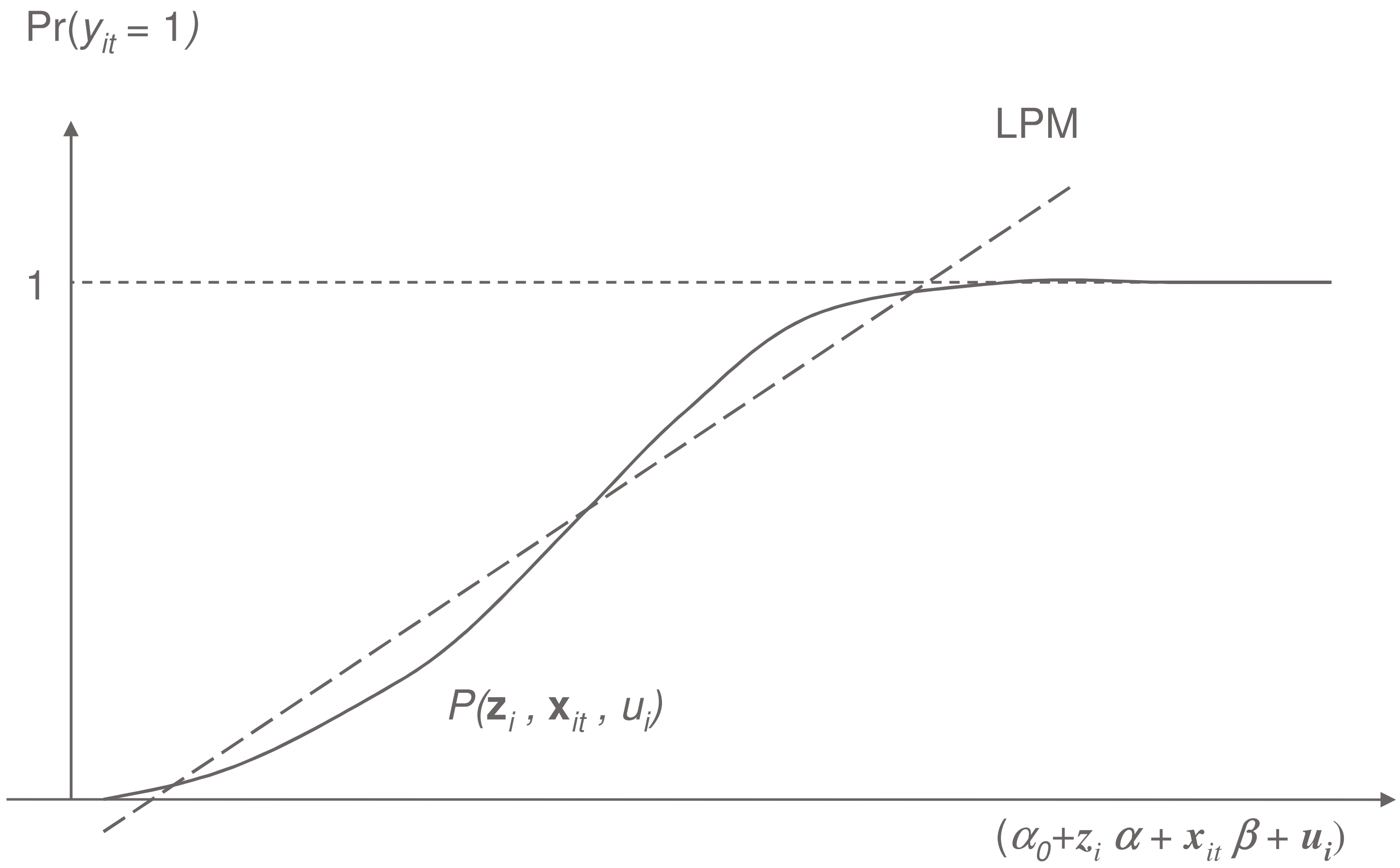
With panel data methods (*e.g.* within-group or random-effects), linear model implies:

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

Disadvantages of the LPM

- Predicted probabilities don't necessarily lie within the 0 to 1 range
- We get a very specific form of heteroskedasticity errors for this model (values are along the continuous OLS line, but Y_i values jump between 0 and 1 - this creates large variation in errors)
- Errors are non-normal

Why nonlinear models are needed



What to do about this?

- To overcome the disadvantages of the LPM, use non-linear methods.
- There are two types of similar S-curves used to analyze these data, *logit* and *probit*
- The two tend to yield similar results
- By fitting a "sigmoidal" or S-shaped, curved line to the data (see chart on left), we can do a much better job of minimizing the errors.

Latent regression models: the binary case

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

Example from global bilateral investment holdings:

If $y_{it}=1$ defines positive investment holdings between two countries , then:

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 invetsment is observed according to:

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

Latent regression models: the binary case cont

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

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

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

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

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

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

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

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

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).

Since the ε_{it} are independent, the joint probability of observing (y_{i1}, $y_{i1}, \ldots, y_{iT_i}$) conditional on u_i (and $\mathbf{z}_i$, $\mathbf{x}_{it}$) is just the product of the conditional probabilities for each time period:

$$\Pr(y_{i1}, \ldots, y_{iT} \mid u_i) = \Pr(y_{i1} \mid u_i) \times \ldots \times \Pr(y_{iT} \mid u_i)$$

$$= F(\alpha_0 + \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{i1} \boldsymbol{\beta} + u_i) \times \ldots \times F(\alpha_0 + \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{iT} \boldsymbol{\beta} + u_i)$$

Random effects logit/probit

Make an assumption about the distribution of u_i (usually assumed to be N(0, σ_u^2)).

Average out *(marginalise with respect to)* the unobservable u_i to get the unconditional probability of the data for individual *i* :

$$\Pr(y_{i1}, \ldots, y_{iT}) = \mathrm{E}[\Pr(y_{i1}, \ldots, y_{iT} \mid u_i)]$$

where "E[.]" refers to the expectation or mean with respect to the N(0, σ_u^2) distribution of u_i .

This unconditional probability $\Pr(y_{i1}, \ldots, y_{iT})$ is the likelihood for individual *i*. Repeat this for all individuals in the sample.

We then choose as our ML estimates the parameter values that maximise the likelihood over the whole sample. This is implemented in Stata, but computing run times are quite long.

This ML method works well only if $\mathrm{cov}(u_i, [\mathbf{z}_i, \mathbf{x}_{it}]) = 0$

Example: Gravity setup (sender-host countries); paired id

- *Investment holdings depend on: the costs, and the attractiveness of the host*
- *xtprobit inv4 logdist comlang_off laginvtreaty lagpolrisk lagfinrisk legor_uk legor_fr legor_ge legor_sc legor_so lagdomcred lagstockturn year1 - year7*

```
Random-effects probit regression               Number of obs      =     15392
Group variable: id_pair                        Number of groups   =      3813

Random effects u_i ~ Gaussian                  Obs per group: min =         1
                                                              avg =       4.0
                                                              max =         6

Integration method: mvaghermite                Integration points =        12

                                               Wald chi2(15)      =   1512.61
Log likelihood  = -4915.4267                   Prob > chi2        =    0.0000
```

inv4	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
logdist	-1.226304	.0795409	-15.42	0.000	-1.382201	-1.070407
comlang_off	1.803748	.2380528	7.58	0.000	1.337173	2.270323
laginvtreaty	.7187815	.1320455	5.44	0.000	.459977	.9775859
lagpolrisk	.0735263	.0065495	11.23	0.000	.0606894	.0863631
lagfinrisk	.0024246	.010589	0.23	0.819	-.0183294	.0231786
legor_uk	-.0489695	.3145673	-0.16	0.876	-.6655102	.5675711
legor_fr	.3420482	.3037728	1.13	0.260	-.2533356	.9374319
legor_ge	.0454518	.3078679	0.15	0.883	-.5579582	.6488619
legor_sc	0	(omitted)				
legor_so	0	(omitted)				
lagdomcred	.0223361	.0014295	15.63	0.000	.0195344	.0251378
lagstockturn	.0053198	.0007708	6.90	0.000	.0038091	.0068305
year1	0	(omitted)				
year2	-.8390752	.0873498	-9.61	0.000	-1.010278	-.6678727
year3	-.6708261	.0835393	-8.03	0.000	-.8345601	-.5070921
year4	-.6179186	.081747	-7.56	0.000	-.7781397	-.4576974
year5	-.152654	.0838404	-1.82	0.069	-.3169782	.0116702
year6	-.2834496	.0716033	-3.96	0.000	-.4237895	-.1431097
year7	0	(omitted)				
_cons	3.693808	.9333702	3.96	0.000	1.864436	5.52318
/lnsig2u	2.617384	.0629678			2.49397	2.740799
sigma_u	3.70133	.1165323			3.479835	3.936923
rho	.9319721	.0039922			.923718	.9393916

```
Likelihood-ratio test of rho=0: chibar2(01) =  6286.51 Prob >= chibar2 = 0.000
```

Estimating output

- In probit and Logit models, interpretation of the parameters is not straightforward
- Not a linear model, so coefficients are not the slope of a line.
- Therefore, if say β_1 is positive (negative), an increase in x would increase (decrease) the probability that the positive outcome would be observed
- Thus, the sign of the estimated parameters tell us if the probability of a +ve outcome will increase or decrease.
- "by how much" the probability increases or decreases is answered by computing the **marginal effects**

Predicted probabilities and marginal effects

- Calculate the predicted (fitted) probability of positive investment holdings
- **Marginal Effect** is 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 is a popular way to present results, partly because the effects can be calculated directly using a standard formula. Can also use Stata command `mfx`.

cont

- Scaling up the effect due to an infinitesimal change is fine in linear models, but not generally 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!).
- 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).
- `mfx` recognises dummy variables and calculates effect due to 0 to 1 change. But `mfx` calculates marginal effect based on infinitesimal change for all other variables (including discrete variables with >2 categories).

```
. mfx, predict(pu0)

Marginal effects after xtprobit
      y  = Pr(inv4=1 assuming u_i=0) (predict, pu0)
         =  .84135461
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
logdist	-.2967176	.02939	-10.10	0.000	-.35432	-.239115	8.4978
comlan~f*	.2040849	.02099	9.72	0.000	.162938	.245232	.104535
laginv~y*	.1682984	.03051	5.52	0.000	.108494	.228103	.453482
lagpol~k	.0177905	.00201	8.84	0.000	.013845	.021736	73.7854
lagfin~k	.0005867	.00256	0.23	0.819	-.004435	.005609	38.4687
legor_uk*	-.0119438	.0773	-0.15	0.877	-.163458	.13957	.336019
legor_fr*	.0800041	.0689	1.16	0.246	-.055029	.215037	.401637
legor_ge*	.0108461	.07247	0.15	0.881	-.131186	.152878	.196985
lagdom~d	.0054045	.00051	10.52	0.000	.004398	.006411	73.905
lagsto~n	.0012872	.00021	6.28	0.000	.000885	.001689	60.2506
year2*	-.2573901	.03337	-7.71	0.000	-.322804	-.191977	.147414
year3*	-.1973371	.03007	-6.56	0.000	-.256273	-.138401	.158134
year4*	-.1792241	.02874	-6.24	0.000	-.235548	-.122901	.161577
year5*	-.0389437	.02265	-1.72	0.086	-.083341	.005454	.142931
year6*	-.0744384	.02074	-3.59	0.000	-.115079	-.033798	.195946

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Presenting results from binary response models

- Marginal effects are often evaluated at mean **x** and **z**, with individual effects set to zero. But:
  - This represents a synthetic, hybrid person that doesn't exist.
  - Technically, no-one has a zero individual effect (prob is zero)
- A more flexible way to present results is to predict probabilities for different combinations of **x** and **z,** representing different types of person or counterfactual scenarios.
- Can present raw probabilities or differences between them (marginal effects). This method can also show the effect of changing any combination of x and z variables simultaneously.

```
. mfx if  comlang_off==1, predict(pu0)

Marginal effects after xtprobit
      y  = Pr(inv4=1 assuming u_i=0) (predict, pu0)
         =   .9948639
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
 logdist |  -.0181612      .01018   -1.78   0.074  -.038118  .001795   8.44694
comlan~f*|   .2176608      .02606    8.35   0.000   .166578  .268743         1
laginv~y*|   .0088611      .00525    1.69   0.092  -.001434  .019156   .335612
lagpol~k |   .0010889      .00062    1.75   0.081  -.000133  .002311   72.6001
lagfin~k |   .0000359      .00016    0.23   0.820  -.000274  .000346   38.2101
legor_uk*|   -.000717      .00456   -0.16   0.875  -.009653  .008219   .594779
legor_fr*|   .0045587      .00457    1.00   0.319  -.004399  .013516   .349285
legor_ge*|   .0006388      .00412    0.16   0.877  -.007438  .008715   .048477
lagdom~d |   .0003308      .00019    1.76   0.078  -.000037  .000699   79.6789
lagsto~n |   .0000788      .00005    1.72   0.086  -.000011  .000169   53.3244
   year2*|  -.0288953      .01448   -2.00   0.046  -.057269 -.000522   .139838
   year3*|  -.0189697      .01003   -1.89   0.059   -.03863   .00069   .155376
   year4*|  -.0163806      .00882   -1.86   0.063  -.033658  .000897   .163456
   year5*|  -.0026079      .00216   -1.21   0.228  -.006845  .001629    .14481
   year6*|   -.005306       .0033   -1.61   0.108  -.011772   .00116   .196395
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

. mfx if  comlang_off==0, predict(pu0)

Marginal effects after xtprobit
      y  = Pr(inv4=1 assuming u_i=0) (predict, pu0)
         =  .79308494
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
 logdist |  -.3503514      .03284  -10.67   0.000   -.41472 -.285983   8.50374
comlan~f*|   .2025304      .02092    9.68   0.000   .161523  .243538         0
laginv~y*|   .2000934      .03571    5.60   0.000     .1301  .270087   .467242
lagpol~k |   .0210062      .00225    9.33   0.000   .016595  .025418   73.9237
lagfin~k |   .0006927      .00303    0.23   0.819  -.005237  .006622   38.4989
legor_uk*|  -.0140985      .09123   -0.15   0.877  -.192908  .164711   .305812
legor_fr*|   .0950568      .08215    1.16   0.247  -.065953  .256067   .407749
legor_ge*|    .012847       .0861    0.15   0.881  -.155908  .181602   .214322
lagdom~d |   .0063814      .00056   11.33   0.000   .005278  .007485    73.231
lagsto~n |   .0015199      .00024    6.43   0.000   .001057  .001983   61.0592
   year2*|  -.2859705      .03451   -8.29   0.000  -.353609 -.218332   .148299
   year3*|  -.2223907      .03203   -6.94   0.000  -.285165 -.159617   .158456
   year4*|  -.2028822      .03088   -6.57   0.000  -.263415  -.14235   .161358
   year5*|  -.0455143      .02618   -1.74   0.082  -.096824  .005795   .142712
   year6*|  -.0864335      .02355   -3.67   0.000  -.132595 -.040272   .195893
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Presenting results from binary response models cont

- It is also possible to calculate
- *average partial effects* (APE) which allow for the average effect of the unobserved individual effects,
- p_1-p_0
- and
- Predicted Probability Ratio: p_1 / p_0

Ordered response models

Ordered response models

- Ordered (or ordinal) variables take discrete values which have a natural ordering:
  - Bank's Capital Adequacy
  - Credit rating (AAA, AA, A, …,CCC)
  - Firm access to credit (deteriorated, unchanged, improved)
- Variables are ordinal but not (necessarily) cardinal, i.e. the "distance" between two categories has no meaning in the model. Only order matters.

Bank's Capital Adequacy according to the Federal Deposit Insurance Corporation

- ***Critically Undercapitalized*** if CAR<2%
- ***Significantly Undercapitalized*** if 2<=CAR<6%
- ***Undercapitalized*** if 6<=CAR<8
- ***Adequately Capitalized*** if 8<=CAR<10%
- ***Well Capitalized*** if CAR>=10% .

- Transform CAR as taking the values: 0, 1, 2, 3, 4

Latent regression (1)

- As in binary response models, assume there is an underlying latent variable y_{it}^* determined as follows:

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

- u_i is assumed to be a random effect distributed independently of $(\mathbf{z}_i, \mathbf{X}_i)$ as N(0, σ_u^2).
- Note there is no constant (see later).
- The observed value of y_{it} is {0, 1, …, J}, depending on where y_{it}^* falls relative to a set of J *cutpoints* or *thresholds*, $\mu_1 < \mu_2 < \ldots < \mu_J$.

Latent regression (2)

- The outcome y_{it} is given as:

$$
\begin{aligned}
y_{it} &= 0 && \text{if } y_{it}^{*} \leq \mu_1 \\
y_{it} &= 1 && \text{if } \mu_1 < y_{it}^{*} \leq \mu_2 \\
& && . \\
y_{it} &= J && \text{if } \mu_J < y_{it}^{*}
\end{aligned}
$$

- So, if $J = 3$, there are 2 cutpoints, μ_1 and μ_2.
- And if $J = 2$ (binary choice model), there is only one cutpoint, μ_1.

Random effects ordered probit (1)

- Assume ε_{it} is normally distributed with unit variance.

$$\Pr(y_{it} = 0 \mid \mathbf{z}_i, \mathbf{x}_{it}, u_i) = \Pr(y_{it}^* \leq \mu_1 \mid \mathbf{z}_i, \mathbf{x}_{it}, u_i)$$
$$= \Pr(\mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + u_i + \varepsilon_{it} \leq \mu_1)$$
$$= \Phi(\mu_1 - \mathbf{z}_i \boldsymbol{\alpha} - \mathbf{x}_{it} \boldsymbol{\beta} - u_i)$$

$$\Pr(y_{it} = 1 \mid \mathbf{z}_i, \mathbf{x}_{it}, u_i) = \Pr(\mu_1 < y_{it}^* \leq \mu_2 \mid \mathbf{z}_i, \mathbf{x}_{it}, u_i)$$
$$= \Pr(\mu_1 < \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + u_i + \varepsilon_{it} \leq \mu_2)$$
$$= \Phi(\mu_2 - \mathbf{z}_i \boldsymbol{\alpha} - \mathbf{x}_{it} \boldsymbol{\beta} - u_i) - \Phi(\mu_1 - \mathbf{z}_i \boldsymbol{\alpha} - \mathbf{x}_{it} \boldsymbol{\beta} - u_i)$$

[which is just Pr $(y_{it}^* \leq \mu_2)$ minus Pr $(y_{it}^* \leq \mu_1)$]

Etc…

Random effects ordered probit (2)

- Finally:

$$\begin{aligned}\Pr(y_{it} = J \mid \mathbf{z}_i, \mathbf{x}_{it}, u_i) &= \Pr(\mu_J < y_{it}^* \mid \mathbf{z}_i, \mathbf{x}_{it}, u_i) \\ &= 1 - \Pr(y_{it}^* \leq \mu_J \mid \mathbf{z}_i, \mathbf{x}_{it}, u_i) \\ &= 1 - \Phi(\mu_J - \mathbf{z}_i \boldsymbol{\alpha} - \mathbf{x}_{it} \boldsymbol{\beta} - u_i)\end{aligned}$$

- Check that these probabilities sum to one!
- Predicting probabilities and calculating marginal effects is done analogously to the binary RE probit.

Random effects ordered probit estimation example (xtoprobit)

```
Random-effects ordered probit regression        Number of obs      =     68125
Group variable: id                              Number of groups   =     10131

Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       6.7
                                                               max =         8

Integration method: mvaghermite                 Integration points =        12

                                                Wald chi2(3)       =   1000.87
Log likelihood  = -5725.3604                    Prob > chi2        =    0.0000
```

cap	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
noi	.277461	.0091586	30.30	0.000	.2595104	.2954115
risk1	-.0134609	.0009823	-13.70	0.000	-.0153862	-.0115356
size1	-.1494287	.016144	-9.26	0.000	-.1810703	-.1177872
/cut1	-7.832985	.2770601	-28.27	0.000	-8.376013	-7.289957
/cut2	-7.009726	.2600338	-26.96	0.000	-7.519383	-6.500069
/cut3	-6.530671	.2533777	-25.77	0.000	-7.027283	-6.03406
/cut4	-5.726291	.2447736	-23.39	0.000	-6.206039	-5.246544
/sigma2_u	.9987348	.10055			.8198859	1.216598

LR test vs. oprobit regression: chibar2(01) = 620.84 Prob>=chibar2 = 0.0000

Obtain predicted probabilities: predict prob*, pu0

```
. sum prob1 prob2 prob3 prob4 prob5
```

Variable	Obs	Mean	Std. Dev.	Min	Max
prob1	68125	.0004656	.0176663	0	1
prob2	68125	.0003956	.0078075	0	.3193535
prob3	68125	.0005367	.0062296	0	.1889938
prob4	68125	.0029317	.0157829	0	.3124515
prob5	68125	.9956704	.0362212	0	1