THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

EFMD EQUIS ACCREDITED

ECON2271

Business Econometrics

Or: Practical Econometrics for Beginners

Week 9: Models for Binary Dependent Variables

Topic 4: Non-continuous Dependent Vars
Agenda and learning outcomes

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

Topic 4: Models for non-continuous dependent variables:

Part (i): Models for binary dependent variables

a) The linear probability model:
   - Understand how this model works, and what its limitations are.
b) The logit and probit models:
   - Understand how these models work, why we use them, and how to interpret model estimates

❖ Text reference: Studenmund Ch. 12 (Gujarati Ch 15 are more detailed)

Part (ii): Other types of non-continuous dependent variables:

c) Models for ordered dependent variables (ordered probit/logit)
d) Models for count data (Poisson models)
e) Models for censored or truncated data (Tobit model)

Topic 4(i): Binary Dependent Variables
a) The linear probability model

- So far, we have worked with many types of model *regressors* (explanatory variables):
  - Continuous and unlimited variables, like wealth (can be positive or negative, with no natural limit)
    - Include as single model regressor (unless you suspect nonlinearities)
  - Continuous and limited variables, like age (limited in at least one direction)
    - Include as single model regressor (unless you suspect nonlinearities)
  - Discrete variables, like the number of children in the household
    - Include as single model regressor (unless you suspect nonlinearities)
  - Ordered variables, like highest qualification obtained
    - Include as a set of dummy variables (*m*-1 dummies for *m* categories)
  - Nominal variables, like marital status (married, separated, divorced, widowed, never married)
    - Include as set of dummy variables (*m*-1 dummies for *m* categories)
  - Binary variables, like gender
    - Include as single dummy variable.
- However, so far, we have always treated our dependent variable as unlimited, continuous, and normally distributed.
- If the dependent variable is not unlimited, continuous or normally distributed, standard OLS estimation may not yield estimates that are BLUE.
  - ➢ Alternative models and estimators may perform better…

Topic 4(i): Binary Dependent Variables
a) The linear probability model

- **Estimating models with binary dependent variables:**
  - Consider what happens if Y is binary:
  - ➢ E.g: *LFP* = indicator for labour force participation for females (1=participant; 0 = non-participant)
    S = Years of schooling; *M* = married; *K* = young children in the household
  - If we want to estimate the effect of *S* on *LFP,* holding *M* and *K* constant, we can estimate the model:

$$LFP_i = b_0 + b_1 S_i + b_2 M_i + b_3 K_i + u_i$$

  - This is a linear probability model:
    - ➢ *Linear* because the right-hand-side of the equation is linear
    - ➢ *Probability* model because the left-hand-side is binary (0/1)
    - ➢ The estimated b_1 coefficient measures the effect of a one-unit change in *S* on *LFP* (nothing new..)
    - ➢ Because *LFP* is binary, we can interpret b_1 as the effect of a one-unit change in *S* on the probability of observing *LFP* = 1.

5

Topic 4(i): Binary Dependent Variables

a) The linear probability model


```
. reg LFP schooling marrdef kids if female==1

      Source |       SS           df       MS      Number of obs   =     5,832
-------------+----------------------------------   F(3, 5828)      =    136.31
       Model |  77.1771727         3  25.7257242   Prob > F        =    0.0000
    Residual |  1099.89004     5,828  .188725127   R-squared       =    0.0656
-------------+----------------------------------   Adj R-squared   =    0.0651
       Total |  1177.06722     5,831  .201863697   Root MSE        =    .43443

------------------------------------------------------------------------------
         LFP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   schooling |   .0457902    .002312    19.81   0.000     .0412579    .0503225
     marrdef |   .0302443   .0136366     2.22   0.027     .0035115    .0569771
        kids |  -.050184   .0118039    -4.25   0.000    -.0733239    -.027044
       _cons |   .0977587   .0329558     2.97   0.003     .0331532    .1623643
------------------------------------------------------------------------------
```


5

Topic 4(i): Binary Dependent Variables

a) The linear probability model

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

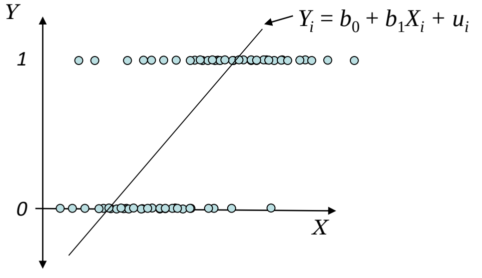
Room for extra notes:

7

Topic 4(i): Binary Dependent Variables
a) The linear probability model

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- **The linear probability model:**



- Clearly, the Y values are lumped in two groups.
- These data are not well spaced around the linear regression line which OLS will produce, and the errors will be strange.
- Even though OLS will probably tell you there is a positive relationship between X and Y, problems include:
  - ➢ Lumpy SEs
  - ➢ R-squared (adjusted or not) will be a poor measure of model fit.
  - ➢ Predicted Y's are continuous will be <0 and >1 for some many values of X.

8

Topic 4(i): Binary Dependent Variables
a) The linear probability model

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- **Solutions for estimating models with binary dependent variables:**

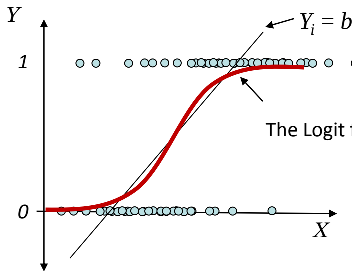
1. Evaluating model fit:
   - Gujarati explains (Ch 15) that the conventionally computed R-squared is of limited value in the LPM model. Values are either 0 or 1, hence no LPM is expected to fit such data well. As a result, the computed R-squared is likely to be much lower than 1; most often between 0.2 and 0.6. We'll only observe high values (>0.8) where the actual data is very closely clustered at the mean values. Then, the regression line will fit well, and predicted Y will be closed to observed Y.
   - Studenmund (Ch 12) suggests calculating a variant called the average $\bar{R}_P^2$ as the proportion of cases correctly predicted by the model. For example, you can make a rule that when $\widehat{LFP}$ > 0.5, this predicts that *LFP* = 1, and that *LFP* = 0 when $\widehat{LFP}$ < 0.5.
   - In the LFP experiment, almost all (99.88 percent) of predictions was LFP=1. Clearly not good, when the average proportion of actual LFP is 72 percent!
   - The proportions of 1s correctly estimated was 99.9% (not surprising…)
   - The proportion of 0s correctly estimated was 0.0024%.. Terrible!
   - This gives an $\bar{R}_P^2$ of about 50 percent. May as well flip a coin?

9

Topic 4(i): Binary Dependent Variables
a) The linear probability model

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- **Solutions for estimating models with binary dependent variables:**



2. Find a better way to estimate the probability of labour force participation: The binomial logit or probit model! First, we consider the logit model:
   - The logit model is S-shaped (sigmoid) and is bounded at 0 and 1.
   - The left-hand-side of the expression is called a logit, and is sometimes expressed as the "log of the odds".
   - Estimation method is different, and so is interpretation of estimated coefficients…

CRICOS Provider Code: 00126G

10

Topic 4(i): Binary Dependent Variables
a) The linear probability model

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- **The logit model:** $\ln\left(\frac{Y_i}{1-Y_i}\right) = b_0 + b_1X_i + u_i$

  - Logits cannot be estimated by OLS, so we another estimator called Maximum Likelihood (ML). ML is an iterative estimation method: *it chooses coefficient estimates that maximise the likelihood of the sample data set being observed.*
  - Actually, if all the classical assumptions behind OLS hold, ML yields identical estimates.
  - Logit models require large sample sizes to be any good (>500 observations recommended).
  - The interpretation of estimated model coefficients are different:
    - b_1 = the change in the log of the odds that Y = 1 when X changes by 1 unit.
    - This is a nonlinear model: The slope changes as X changes
    - One common approach is therefore to calculate Y at mean X values.
    - We can also calculate marginal effects by using partial derivatives: the change in Y with respect to X = $b_1[Y_i(1 - Y_i)]$
    - You can use a rough estimate of 0.25: Multiply b1 by 0.25, and you get an equivalent linear probability model coefficient.
  - In Stata, the command for logit estimation is *logit* and the default option actually produces marginal effects. The additional option of [, *or*] after the *logit* command (or a *logistic* command) will produce odds ratios. However, Stata can also calculate marginal effects at mean values post estimation – directly comparable to LPM estimates and very useful!

11

Topic 4(i): Binary Dependent Variables

a) The linear probability model


```
. logit LFP schooling marrdef kids if female==1

Iteration 0:   log likelihood =  -3460.971
Iteration 1:   log likelihood = -3267.0065
Iteration 2:   log likelihood = -3263.6317
Iteration 3:   log likelihood = -3263.6284
Iteration 4:   log likelihood = -3263.6284

Logistic regression                             Number of obs     =      5,832
                                                LR chi2(3)        =     394.69
                                                Prob > chi2       =     0.0000
Log likelihood = -3263.6284                     Pseudo R2         =     0.0570

------------------------------------------------------------------------------
         LFP |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   schooling |    .241369   .0129132    18.69   0.000     .2160595    .2666784
     marrdef |   .1549662   .0710508     2.18   0.029     .0157093    .2942232
        kids |  -.2717887   .0622608    -4.37   0.000    -.3938178   -.1497597
       _cons |  -2.250429   .1751627   -12.85   0.000    -2.593741   -1.907116
------------------------------------------------------------------------------
```

- Stata produces a Pseudo-R-squared for models estimated by ML, which compared the estimated model against an intercept-only model.

12

Topic 4(i): Binary Dependent Variables

a) The linear probability model

The University of Western Australia
Achieve International Excellence

- Stata can also calculate the probability of a female being in the labour force, at mean values of the explanatory variables.

```
. margins, atmeans

Adjusted predictions                            Number of obs     =      5,832
Model VCE    : OIM

Expression   : Pr(LFP), predict()
at           : schooling       =    13.50969 (mean)
               marrdef         =    .7654321 (mean)
               kids            =    .3993484 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .7350638   .0060808   120.88   0.000     .7231457     .746982
------------------------------------------------------------------------------
```

13

Topic 4(i): Binary Dependent Variables
a) The linear probability model

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- You can specify the variable values for those you are concerned about, and Stata will treat the rest as at their means…

```
. margins, at(marrdef=1 kids==1) atmeans

Adjusted predictions                            Number of obs     =      5,832
Model VCE    : OIM

Expression   : Pr(LFP), predict()
at           : schooling       =    13.50969 (mean)
               marrdef         =           1
               kids            =           1

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |    .709626   .0099659    71.21   0.000     .6900932    .7291588
------------------------------------------------------------------------------
```


14

Topic 4(i): Binary Dependent Variables
a) The linear probability model

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- You can then predict the difference in the probability of being in the labour force for a female with kids versus no kids, at mean years of schooling:

```
. margins, at(marrdef=1 kids==0) atmeans

Adjusted predictions                            Number of obs     =      5,832
Model VCE    : OIM

Expression   : Pr(LFP), predict()
at           : schooling       =    13.50969 (mean)
               marrdef         =           1
               kids            =           0

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .7623047   .0083486    91.31   0.000     .7459419    .7786676
------------------------------------------------------------------------------
```

15

Topic 4(i): Binary Dependent Variables

a) The linear probability model

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- Even more useful is a command which calculates marginal effects for a change in X, at mean values:

```
. margins, dydx(*) atmeans post

Conditional marginal effects                    Number of obs     =      5,832
Model VCE    : OIM

Expression   : Pr(LFP), predict()
dy/dx w.r.t. : schooling marrdef kids
at           : schooling       =    13.50969 (mean)
               marrdef         =    .7654321 (mean)
               kids            =    .3993484 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   schooling |   .0470054   .0024332    19.32   0.000     .0422363    .0517745
     marrdef |   .0301789   .0138325     2.18   0.029     .0030677    .0572901
        kids |  -.0529295   .0121033    -4.37   0.000    -.0766515   -.0292075
------------------------------------------------------------------------------
```


16

Topic 4(i): Binary Dependent Variables

a) The linear probability model

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

(room for notes)

17

Topic 4(i): Binary Dependent Variables
a) The linear probability model

- **The probit model:** $Z_i = \Phi^{-1}(P_i) = b_0 + b_1 X_i + u_i$
  - $\Phi^{-1}(P_i)$ is the inverse of the normal cumulative distribution function.
  - The probit function has very similar attributes to the logit function: it has very similar shape, is estimated by ML, and has the same requirement wrt sample size and measurement of goodness-of-fit.
  - But the estimated model coefficients have a different scale (one logit coeff ≈ 1.6 probit coeff)
    - But you can still use that handy post-estimation command [*margins, dydx(*) atmeans post*], in which case you don't have to worry about it.

18

Topic 4(i): Binary Dependent Variables
a) The linear probability model

```
. probit LFP schooling marrdef kids if female==1

Iteration 0:   log likelihood =  -3460.971
Iteration 1:   log likelihood = -3265.2218
Iteration 2:   log likelihood = -3264.2878
Iteration 3:   log likelihood = -3264.2877

Probit regression                               Number of obs     =      5,832
                                                LR chi2(3)        =     393.37
                                                Prob > chi2       =     0.0000
Log likelihood = -3264.2877                     Pseudo R2         =     0.0568

------------------------------------------------------------------------------
         LFP |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   schooling |   .1425338    .007492    19.02   0.000     .1278497     .157218
     marrdef |   .0897337   .0425837     2.11   0.035     .0062711    .1731963
        kids |  -.1636176   .0370247    -4.42   0.000    -.2361846   -.0910506
       _cons |   -1.31063   .1035069   -12.66   0.000      -1.5135    -1.10776
------------------------------------------------------------------------------
```

Topic 4(i): Binary Dependent Variables

a) The linear probability model

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

```
. margins, dydx(*) atmeans post

Conditional marginal effects                    Number of obs     =      5,832
Model VCE    : OIM

Expression   : Pr(LFP), predict()
dy/dx w.r.t. : schooling marrdef kids
at           : schooling       =    13.50969 (mean)
               marrdef         =    .7654321 (mean)
               kids            =    .3993484 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   schooling |   .0469692   .0024425    19.23   0.000      .042182    .0517564
     marrdef |   .0295699   .0140324     2.11   0.035      .002067    .0570729
        kids |  -.0539169   .0121906    -4.42   0.000      -.07781   -.0300238
------------------------------------------------------------------------------
```

CRICOS Provider Code: 00126G

Topic 4(i): Binary Dependent Variables

a) The linear probability model

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- **How to choose the best model?**
  - Comparing the LPM and logit models presented here, the logit model did less badly in the sense that more of the 0s were correctly predicted (56 of them, rather than just 4. So the model is slightly better at finding the 0s. However, the $\bar{R}_P^2$ statistic is now only increased to 0.51..
  - Should you choose the logit or probit model? Shouldn't matter – results are likely to be very similar. If in doubt, do both so you can check the robustness of your results. The shapes of the underlying functions are very similar, but subtly different – mostly at the tails (see below). Hence, it might matter in some applications.

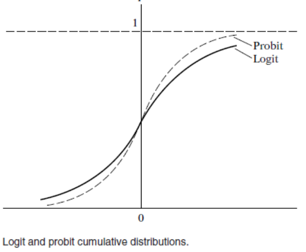


FIGURE 15.6 Logit and probit cumulative distributions.

CRICOS Provider Code: 00126G

21

Topic 4(i): Binary Dependent Variables
a) The linear probability model

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- **Let's do another one: estimating the probability of having kids (for women)**

```
. reg kids lninc lnwages lnwealth schooling age agesq if female==1

      Source |       SS           df       MS      Number of obs   =     3,770
-------------+----------------------------------   F(6, 3763)      =    194.86
       Model |  211.081664         6  35.1802773   Prob > F        =    0.0000
    Residual |  679.362899     3,763  .180537576   R-squared       =    0.2371
-------------+----------------------------------   Adj R-squared   =    0.2358
       Total |  890.444562     3,769  .236254859   Root MSE        =     .4249

------------------------------------------------------------------------------
        kids |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lninc |   .0444408   .0122705     3.62   0.000     .0203833    .0684983
     lnwages |  -.0777184   .0071685   -10.84   0.000    -.0917729   -.0636638
    lnwealth |    .016425   .0060605     2.71   0.007     .0045429    .0283071
   schooling |   .0010717   .0030653     0.35   0.727    -.004938     .0070814
         age |   .0994679   .0056103    17.73   0.000     .0884683    .1104674
       agesq |  -.0013141    .000063   -20.86   0.000    -.0014376   -.0011906
       _cons |  -1.218137   .1680056    -7.25   0.000    -1.547528   -.8887458
------------------------------------------------------------------------------
```


22

Topic 4(i): Binary Dependent Variables
a) The linear probability model

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- **Let's do another one: estimating the probability of having kids (for women)**

```
. logit kids lninc lnwages lnwealth schooling age agesq if female==1

Iteration 0:   log likelihood = -2508.0738
Iteration 1:   log likelihood = -1911.4224
Iteration 2:   log likelihood = -1823.9753
Iteration 3:   log likelihood =  -1820.538
Iteration 4:   log likelihood = -1820.5267
Iteration 5:   log likelihood = -1820.5267

Logistic regression                             Number of obs     =      3,770
                                                LR chi2(6)        =    1375.09
                                                Prob > chi2       =     0.0000
Log likelihood = -1820.5267                     Pseudo R2         =     0.2741

------------------------------------------------------------------------------
        kids |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lninc |    .455405   .0958004     4.75   0.000     .2676396    .6431704
     lnwages |  -.5345653   .0506023   -10.56   0.000     -.633744   -.4353865
    lnwealth |   .0663709   .0373191     1.78   0.075    -.0067733     .139515
   schooling |    -.00588   .0183588    -0.32   0.749    -.0418625    .0301025
         age |   1.101115   .0515112    21.38   0.000     1.000155    1.202075
       agesq |  -.0144092   .0006461   -22.30   0.000    -.0156755    -.013143
       _cons |  -20.47227   1.339051   -15.29   0.000    -23.09676   -17.84778
------------------------------------------------------------------------------
```

CRICOS Provider Code: 00126G

Topic 4(i): Binary Dependent Variables

a) The linear probability model

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- **Let's do another one: estimating the probability of having kids (for women)**

```
. margins, dydx(*) atmeans post

Conditional marginal effects                    Number of obs     =      3,770
Model VCE    : OIM

Expression   : Pr(kids), predict()
dy/dx w.r.t. : lninc lnwages lnwealth schooling age agesq
at           : lninc           =     11.51531 (mean)
               lnwages         =     10.56226 (mean)
               lnwealth        =     12.86391 (mean)
               schooling       =     13.92679 (mean)
               age             =     43.56764 (mean)
               agesq           =     2016.641 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lninc |   .0815469   .0170246     4.79   0.000     .0481794    .1149144
     lnwages |  -.0957217   .0092877   -10.31   0.000    -.1139253   -.0775181
    lnwealth |   .0118847      .0067     1.77   0.076     -.001247    .0250164
   schooling |  -.0010529   .0032855    -0.32   0.749    -.0074924    .0053866
         age |   .1971707   .0071061    27.75   0.000      .183243    .2110983
       agesq |  -.0025802   .0000853   -30.23   0.000    -.0027475   -.0024129
------------------------------------------------------------------------------
```