PPOL502-01, Spring 2016
Course Notes # 15: Binary Dependent Variable: Logit and Probit Models

I. FIXING THE DISADVANTAGES OF THE LINEAR PROBABILITY MODEL (LPM)

Since the key disadvantage of the linear probability model is that it produces predicted probabilities that fall outside the valid range of a probability, perhaps there are some easy fixes we can implement.

LPM: $P\ (Y=1|\mathbf{X}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 .... \beta_k X_k$

A simple way is to find a function **F( )** such that $P\ (Y=1|\mathbf{X})$ lies in the valid range of a probability.

So our task is to find an F() such that:

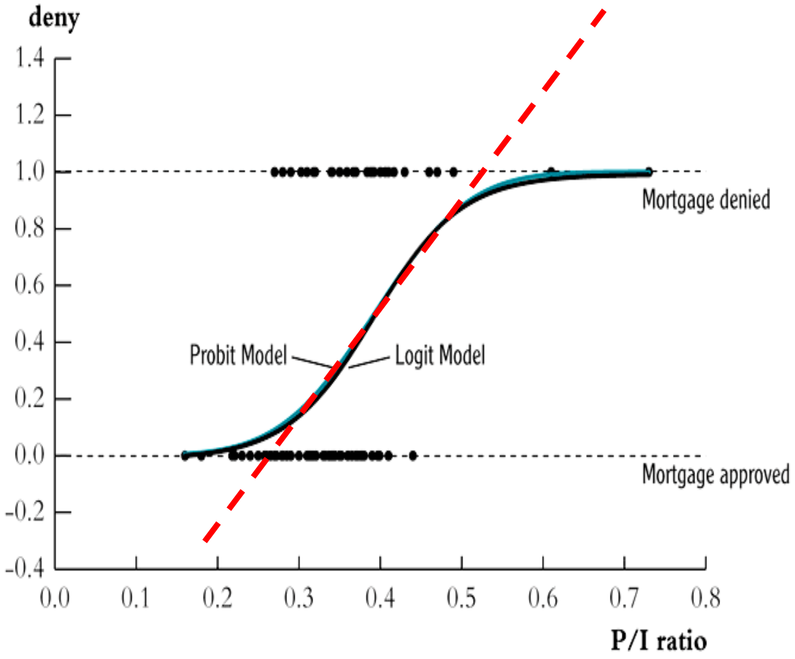
$$P\ (Y=1|\mathbf{X}) = \mathrm{F}(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 .... \beta_k X_k)$$

lies in the valid range of a probability.

The most natural functional forms that satisfy this condition are ***cumulative distribution functions (CDFs)***

FIGURE 9.3 Probit and Logit Models of the Probability of Denial, Given the P/I Ratio

These logit and probit models produce nearly identical estimates of the probability that a mortgage application will be denied, given the payment-to-income ratio.



- The typical choice is to either use the CDF of the normal distribution (probit) or the CDF of the logistic distribution (logit). There is no definitive reason to prefer one to the other. Most economists prefer the normal distribution.

By imposing these CDFs on our model we can now re-write the predicted probability of success as follows:

Probit: $P\ (Y=1|\mathbf{X}) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 .... \beta_k X_k)$

For the **probit**, this turns our initial prediction into a z-score that we can read off our normal distribution table to calculate the predicted probability.

$P\ (Y=1|\mathbf{X}) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 .... \beta_k X_k) = \Phi(z)$,

Where $z = (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 .... \beta_k X_k$

For the Logit regression, the CDF differs but the principle is the same.

Logit: $P\ (Y=1|\mathbf{X}) = F(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 .... \beta_k X_k) = F(z)$

Where $z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 .... \beta_k X_k$

$$F(z) = \frac{1}{1 + e^{-z}}$$

Lets go back to our example of looking at the likelihood of being in the service profession.

The linear probability model predicted the following expected probabilities:

(a) a person with no education, male, who is white = **0.341**

(b) a female with no education, who is white = 0.341 + 0.101 = **0.442**

(c) a female with 10 years of education who is nonwhite =0.341 - (0.019*10) + 0.101 – 0.046 = **0.206**

In other words, our best guess of the percent of the (c) group who work in a service occupation is 20.6%.

(d) a male with 18 years of education who is nonwhite = 0.341 – (0.019*18) – 0.046 = **- 0.047**: an out-of-range prediction!

To see what the corresponding values are for the Normal and Logistic CDFs, we need to run probit and logit models to generate new *betas* and corresponding probabilities.

II. Example using the Normal CDF

$$P(Y=1\,|\,\mathbf{X}) = \frac{1}{\sqrt{2\pi}\int_{-\infty}^{z} e^{-s^2/2}ds}, \text{ where z} = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k.$$

In STATA, the command for the probit is simply **probit**

```
probit servocc  educ female nonwhite
```

```
Iteration 0:   log likelihood = -213.66328
Iteration 1:   log likelihood = -200.05342
Iteration 2:   log likelihood = -199.83821
Iteration 3:   log likelihood = -199.83803
Iteration 4:   log likelihood = -199.83803

Probit regression                                 Number of obs   =        526
                                                  LR chi2(3)      =      27.65
                                                  Prob > chi2     =     0.0000
Log likelihood = -199.83803                       Pseudo R2       =     0.0647

------------------------------------------------------------------------------
     servocc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  -.0945167   .0257061    -3.68   0.000    -.1448997   -.0441338
      female |   .4976532   .1432958     3.47   0.001     .2167985    .7785078
    nonwhite |  -.2379628   .2480437    -0.96   0.337    -.7241196    .2481939
       _cons |  -.1807477   .3283512    -0.55   0.582    -.8243043    .4628089
------------------------------------------------------------------------------
```

The coefficients of this regression **can no longer be interpreted directly**. Instead we need to use them to calculate what the predicted probability is for **given values of the Xs.**

For (a) a person with no education, male, who is white = **0.341 (LPM)**

Using the probit function above we need to calculate

$\Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3) = \Phi(\text{z})$,

For white males with no schooling, $X_1=0, X_2=0, X_3=0,$

So $\Phi(\text{z}) = \Phi(\beta_0) = \Phi(-.1807477) =$ **0.428**

In STATA, we can implement both of the steps above by typing:

```
gen z = _coef[_cons] + _coef[educ]*0 + _coef[female]*0 + _coef[nonwhite]*0
```

This calculates z =-.1807477. We can then use STATA in-built normal tables function **normprob** to calculate the probability associated with this z-score.

```
display normprob(z)       [=0.428]
```

The probit model calculates that 43% of white males work in a service occupation

We can repeat this exercise for:
(b) a female with no education, who is white: Our LPM estimate is **0.442**

In STATA we can calculate the corresponding z-score using the coefficients from the probit regression above by typing:

```
gen z = _coef[_cons] + _coef[educ]*0 + _coef[female]*1 + _coef[nonwhite]*0
```

and then

```
display normprob(z)
```

We obtain a predicted probability for this group of observations of 0.624

Finally for the set of observations for which the LPM predicted a **negative probability,**

(d) a male with 18 years of education who is nonwhite = 0.341 – (0.019*18) – 0.046 = **- 0.047**:

The probit model generates the following outcome;

```
gen z = _coef[_cons] + _coef[educ]*18 + _coef[nonwhite]*1 = -2.120011
display normprob(-2.120011) = 0.01700254
```

Note that for both of these extreme cases (we have very few or no male/female observations with 0 years of schooling), we obtain a deviation between the predictions of the LPM and the Probit regression. In particular the predictions from the probit are consistently higher than the LPM predictions.

Now let's try the same exercise for more typical values of education. Let's compare the predicted probabilities for white males/females with average levels of education (12.56274)

Probit -White Males:

```
gen z = _coef[_cons] + _coef[educ]*12.56274 + _coef[female]*0  =  -1.368

display normprob(-1.368) =    0.085
```

Probit -White Females:

```
gen z = _coef[_cons] + _coef[educ]*12.56274 + _coef[female]*1  =  -0.870

display normprob(-0.87) =     0.192
```

We can obtain corresponding values for the linear probability model by multiplying out our estimated LPM coefficients as before.

LPM -White Males:

0.341 - (0.019*12.56274) = `0.102`

LPM -White Females:

0.341 - (0.019*12.56274) + 0.101 = `0.203`

As you can tell the predicted probabilities of working in a service occupation for both white males (0.085 vs 0.10) and females (0.192 vs 0.203) are now much closer to each other.

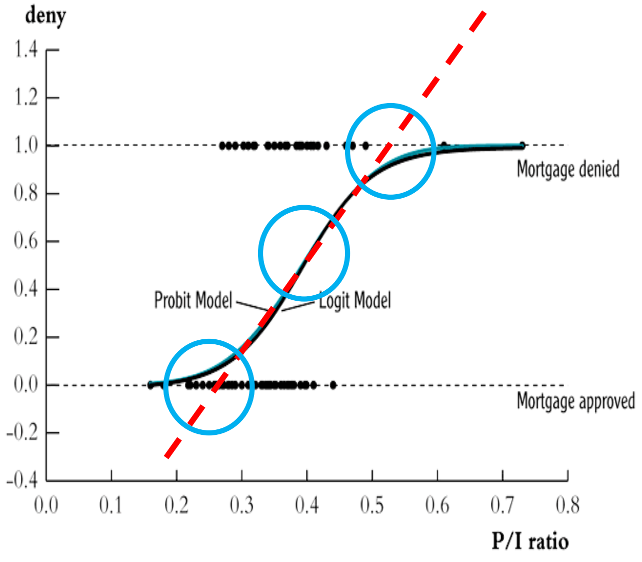
Consequently the effect of being female (holding race and education constant at white and mean of education) is also closer:

The probit estimates suggest that the effect of being female is `0.192 -0.085 = .107`
This is very close to the effect of being female in the LPM `(0.101)`

The differences at the extreme ends of the X's reflect the curvature of the CDF that constrains the probabilities to be well behaved.

FIGURE 9.3 Probit and Logit Models of the Probability of Denial, Given the P/I Ratio

These logit and probit models produce nearly identical estimates of the probability that a mortgage application will be denied, given the payment-to-income ratio.



III. Example using the Logistic CDF

To predict the probabilities of working in a service occupation we can also use the logistic CDF instead of the probit.

$$P\ (Y{=}1|\mathbf{X}) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\cdots+\beta_k X_k)}}$$

In STATA, the command to generate estimates from this CDF is **logit**

```
logit servocc  educ female nonwhite
```

```
Iteration 0:   log likelihood = -213.66328
Iteration 1:   log likelihood = -200.94389
Iteration 2:   log likelihood = -200.28479
Iteration 3:   log likelihood = -200.28234
Iteration 4:   log likelihood = -200.28234

Logistic regression                               Number of obs   =        526
                                                  LR chi2(3)      =      26.76
                                                  Prob > chi2     =     0.0000
Log likelihood = -200.28234                       Pseudo R2       =     0.0626

------------------------------------------------------------------------------
     servocc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  -.1658479   .0467108    -3.55   0.000    -.2573994   -.0742964
      female |   .9181429   .2696946     3.40   0.001     .3895511    1.446735
    nonwhite |  -.4180149   .4644127    -0.90   0.368    -1.328247    .4922174
       _cons |  -.2805028   .5892052    -0.48   0.634    -1.435324    .8743181
------------------------------------------------------------------------------
```

As with the probit regression above, we can calculate the predicted probabilities for white males/females with no/average schooling as follows. The STATA command for obtaining the predicted z-scores is as before. However, since we are using the logistic distribution, we need a 'different' set of tables to obtain our predicted probability.

Logit – White Males with 0 years of Schooling

```
gen z = _coef[_cons] + _coef[educ]*0 + _coef[female]*0 = -.2805

display 1/(1+exp(-z)) = 1/(1+exp(.2805))  = 0.430
```

Logit – White Females with 0 years of Schooling

```
gen z = _coef[_cons] + _coef[educ]*0 + _coef[female]*1 = .6376

display 1/(1+exp(-z)) = 1/(1+exp(.6376))  = 0.654
```

We can repeat this for white male/female observations with mean years of schooling.

Logit – White Males with mean years of Schooling

```
gen z = _coef[_cons] + _coef[educ]*12.56 + _coef[female]*0 =-2.364

display 1/(1+exp(-z)) = 1/(1+exp(2.264)) = 0.086
```

Logit – White Females with mean years of Schooling

```
gen z = _coef[_cons] + _coef[educ]*12.56 + _coef[female]*1 =-1.446

display 1/(1+exp(-z)) = 1/(1+exp(1.446)) = 0.191
```

We can also estimate the predicted probability for the negative predicted probability case that we obtained with the LPM.

Logit – Non-white Males with 18 years of Schooling

```
gen probpred = _coef[_cons] + _coef[educ]*18 + _coef[nonwhite]*1 =-3.684

display 1/(1+exp(-z)) = 1/(1+exp(3.684)) = 0.025 → fatter tails
```

The Table below summarizes our predicted probabilities from using the LPM, probit and logit models.

Case	LPM	Probit	Logit
White Male, 0 schooling	.341	.428	.430
White Female, 0 schooling	.442	.624	.654
Effect of being female	.101	.196	.224
White Male, mean schooling	.102	.085	.086
White Female, mean Schooling	.203	.192	.191
Effect of being female	.101	.107	.105
Non-white male with 18 years of school	-.047	.017	.025

Note again the differences (across models) in the effect of being female at the extreme values of education (0 schooling), but the greater similarities at the mean of schooling.

IV: Interpreting coefficients in Probit and Logit Models

One of the difficulties with probit and logit models is that interpreting coefficient estimates is not straightforward.

Unlike the LPM, where a unit change in X corresponds to a Beta change in the predicted probability of success, the highly non-linear nature of probit/logit models implies that the impact of a marginal change in X on a change in the Probability of Y cannot be inferred easily from the regression output.

Also, since the models are nonlinear, the marginal impact will vary with different levels of X.

There are two ways [that we will look at in this class] of interpreting coefficients from probit/logit models. The first uses the 'prediction' approach used above. The second uses STATA commands to provide coefficient estimates that are more readily interpretable.

Recall the interpretation of the coefficient on education in the LPM:

- β_{EDUC}: This effect is statistically significant (p<0.0001):

  Holding gender and race constant, for each additional year of schooling, the predicted probability of working in a service occupation decreases by 0.019, on average..

  In other words, there is a -1.9 *percentage point* effect of an additional year of education (i.e., almost 2 percentage points), conditional on the other variables in the model.

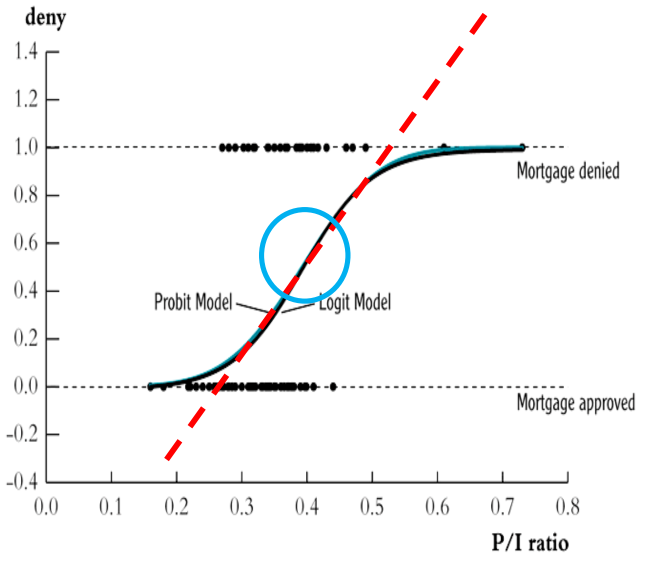
This effect is the same whether we are going from the 10 → 11 years of schooling, or 12.56 -→ 13.56 years of schooling or 15 → 16 years of schooling.

The table below shows the predicted probabilities of working in a service occupation for a *non-white female* with different levels of schooling for each of the three models.

Non-white female with	LPM	Probit	Logit
10 years of schooling	.206	.193	.192
11 years schooling	.187	.168	.167
Effect of an additional year of school	-.019	-.025	-.025
12.56 years of schooling	.157	.134	.134
13.56 years schooling	.138	.115	.116
Effect of an additional year of school	-.019	-.019	-.018
15 years of schooling	.111	.09	.094
16 years schooling	.092	.076	.081
Effect of an additional year of school	-.019	-.014	-.013

This table allows us to report how an additional year of schooling changes the probability of working in a service occupation at 10, 12.56 and 15 years of schooling for each of the models. As the ***probit*** and ***logit*** models' estimates suggest, the effect of an additional year of education varies from -2.5 percentage points at 10 years of schooling to -1.3 percentage points at 15 years of schooling (holding gender (female) and race (non-white) constant). The effects suggested by the ***probit*** and ***logit*** converge to the LPM effect at the mean of schooling.

FIGURE 9.3 Probit and Logit Models of the Probability of Denial, Given the P/I Ratio



These logit and probit models produce nearly identical estimates of the probability that a mortgage application will be denied, given the payment-to-income ratio.

The second interpretation strategy relies on obtaining coefficients from STATA that are more readily interpretable.

For the probit, STATA can report the marginal effect (dF/dx) with respect to all the X's (Like P(Y=1|X) = F(z), this marginal effect has to be estimated for particular values of the X. STATA generally reports dF/dx at the mean of the X's).

The marginal effect is the slope of the tangent to the CDF at the mean of X.

There are two ways to generate marginal effects in STATA. The first is the **dprobit** command which asks STATA to show marginal effects instead of the Betas.

The second is the **mfx compute** command which is run after the **probit** command.

dprobit - to obtain the marginal effects of a probit estimation in STATA, we type:

```
dprobit servocc  educ female nonwhite
```

```
Iteration 0:   log likelihood = -213.66328
Iteration 1:   log likelihood = -200.05342
Iteration 2:   log likelihood = -199.83821
Iteration 3:   log likelihood = -199.83803

Probit regression, reporting marginal effects           Number of obs =    526
                                                        LR chi2(3)    =  27.65
                                                        Prob > chi2   = 0.0000
Log likelihood = -199.83803                             Pseudo R2     = 0.0647

------------------------------------------------------------------------------
 servocc |      dF/dx   Std. Err.      z    P>|z|     x-bar  [    95% C.I.   ]
---------+--------------------------------------------------------------------
    educ |  -.0193717   .0051482    -3.68   0.000   12.5627  -.029462 -.009281
  female*|   .1035413   .0296244     3.47   0.001   .479087   .045478  .161604
nonwhite*|  -.0435916   .0401602    -0.96   0.337   .102662  -.122304  .035121
---------+--------------------------------------------------------------------
  obs. P |   .1406844
 pred. P |     .12422  (at x-bar)
------------------------------------------------------------------------------
```

(*) dF/dx is for discrete change of dummy variable from 0 to 1
z and P>|z| correspond to the test of the underlying coefficient being 0

Note that the marginal effect for indicators considers a discrete change that corresponds to "switching on" the indicator – going from 0 → 1. As you can see the marginal effects reported here are similar to the Betas obtained in the linear probability model.

The interpretation of these marginal effects is as follows:

dF/d$_{EDUC}$: This effect is statistically significant ($p<0.001$):

> Holding gender and race constant at ***their means***, each additional year of schooling, at the mean of schooling, reduces the predicted probability of working in a service occupation by 0.0194, on average.

dF/d$_{FEMALE}$: This effect is statistically significant ($p<0.01$):

> Holding education and race constant at ***their means***, females are about 10.35 percentage points more likely to work in the service occupation than males, on average.

We can also obtain these marginal effects using the **mfx compute** command right after running the probit command. For example, if we want to know what the marginal effect of an additional year of schooling is for a non-white female with 10 years of schooling we can type.

```
probit servocc  educ female nonwhite
mfx compute, at(educ=10, female=1, nonwhite=1)
```

The resulting output is shown below.

```
Marginal effects after probit
      y  = Pr(servocc) (predict)
         =  .19318352
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
    educ |   -.025911      .00932   -2.78   0.005  -.044175 -.007647        10
  female*|   .1068805      .04082    2.62   0.009    .02688  .186881         1
nonwhite*|  -.0717327      .06965   -1.03   0.303  -.208237  .064772         1
------------------------------------------------------------------------------
```

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

As you can tell, the marginal effect at education = 10 is nearly identical to the effect we calculated using the predicted probabilities in going from 10 → 11 years of schooling (**-0.025**). The difference is due to the curvature of the CDF at that point.

```
. mfx compute, at(educ=15, female=1, nonwhite=1)
```

```
Marginal effects after probit
      y  = Pr(servocc) (predict)
         =  .09031657
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
    educ |  -.0153887       .0054   -2.85   0.004  -.025972 -.004806        15
  female*|   .0571718      .02811    2.03   0.042   .002087  .112257         1
nonwhite*|  -.0451654       .0415   -1.09   0.277  -.126513  .036182         1
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Again accounting for some curvature of the CDF, the marginal effect at education =15 is nearly identical to our estimate obtained by looking at the difference in predicted probabilities of working in a service occupation in going from 15 → 16 years of schooling (**-0.014**)

****

We can also ask STATA to provide us with more interpretable coefficients from the logit regression. However, the resulting coefficients are not as easy to interpret as the marginal effects above.

Recall that the logistic CDF implies that we estimate the following:

$$P\ (Y=1|\mathbf{X}) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\cdots+\beta_k X_k)}}$$

A simple re-arrangement of this can produce

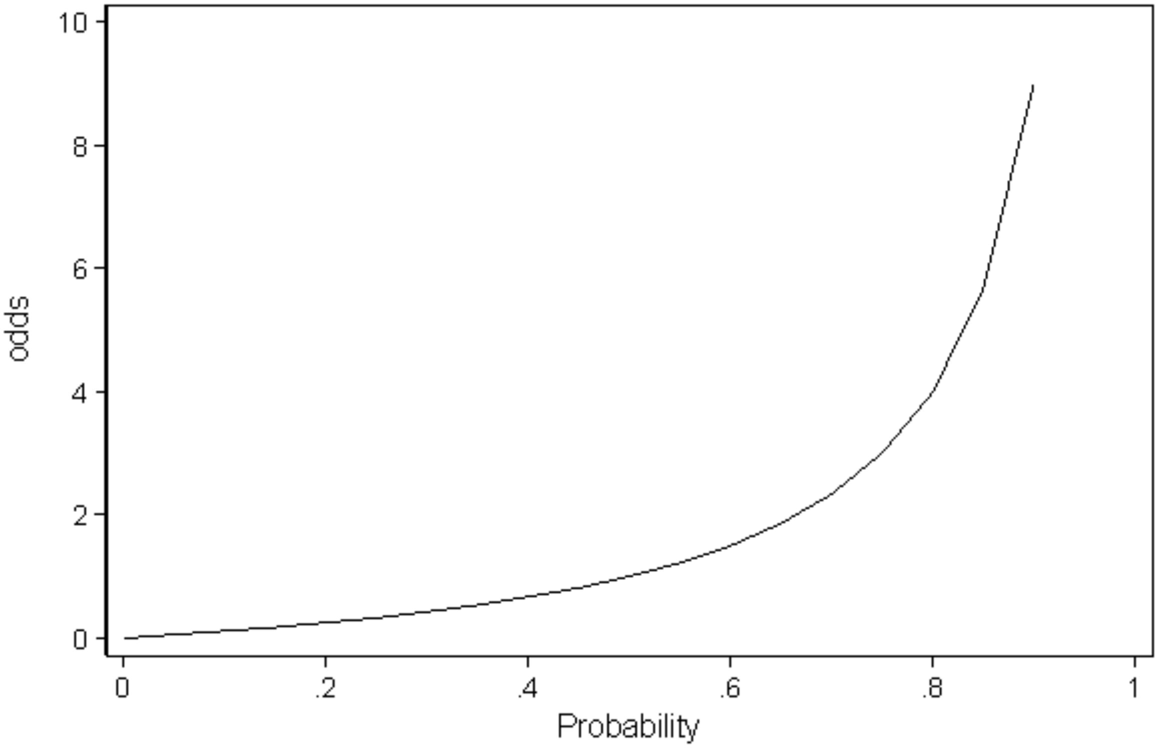
$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 .... \beta_k X_k$$

While this may look like a nice log-level specification, **it is not.**
This is because p = $P\ (Y=1|\mathbf{X})$ is a function of X.

We have now transformed the dependent variable into the log of the **odds ratio [p/(1-p)]**.

If we now exponentiate both sides of this function we obtain.

$$\frac{p}{1-p} = e^{(\beta_0+\beta_1 X_1+\cdots+\beta_k X_k)}$$



We can now ask STATA to report the odd-ratio version of the logit regression – exp(beta1), exp(beta2).

```
logit servocc  educ female nonwhite
Iteration 0:   log likelihood = -213.66328
Iteration 1:   log likelihood = -200.94389
Iteration 2:   log likelihood = -200.28479
Iteration 3:   log likelihood = -200.28234
Iteration 4:   log likelihood = -200.28234

Logistic regression                               Number of obs   =        526
                                                  LR chi2(3)      =      26.76
                                                  Prob > chi2     =     0.0000
Log likelihood = -200.28234                       Pseudo R2       =     0.0626

------------------------------------------------------------------------------
     servocc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  -.1658479   .0467108    -3.55   0.000    -.2573994   -.0742964
      female |   .9181429   .2696946     3.40   0.001     .3895511    1.446735
    nonwhite |  -.4180149   .4644127    -0.90   0.368    -1.328247    .4922174
       _cons |  -.2805028   .5892052    -0.48   0.634    -1.435324    .8743181
------------------------------------------------------------------------------  .
```

```
logit servocc  educ female nonwhite, or
```

```
Iteration 0:   log likelihood = -213.66328
Iteration 1:   log likelihood = -200.94389
Iteration 2:   log likelihood = -200.28479
Iteration 3:   log likelihood = -200.28234
Iteration 4:   log likelihood = -200.28234

Logistic regression                               Number of obs   =        526
                                                  LR chi2(3)      =      26.76
                                                  Prob > chi2     =     0.0000
Log likelihood = -200.28234                       Pseudo R2       =     0.0626

------------------------------------------------------------------------------
     servocc | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .8471751   .0395722    -3.55   0.000     .7730594    .9283964
      female |   2.504635   .6754866     3.40   0.001     1.476318    4.249217
    nonwhite |   .6583524   .3057473    -0.90   0.368     .2649413     1.63594
------------------------------------------------------------------------------
```

The coefficient on education is significant (p-value <0.001).

The coefficient (OR) for education suggests that holding gender and race constant, a one year increase in schooling is associated *increasing the* odds ratio by a factor of 0.847 (exp (-.1658479)) or a 15.3 percent reduction in the odds of working in a service occupation.

The OR coefficient on gender suggests that holding education and race constant, the odds of working in a service occupation are 2.5 times larger for females than the odds of males.

Alternatively, we can say that the odds of working in a service occupation are 150% higher for females (relative to males).

We can also compute these marginal effects using the **mfx compute** command.

```
. mfx compute, at(educ=10, female=1, nonwhite=1)

Marginal effects after logit
      y  = Pr(servocc) (predict)
         =   .1917229
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
    educ |  -.0257007      .01066   -2.41   0.016   -.04659 -.004812        10
  female*|   .1052117      .04484    2.35   0.019    .01732  .193104         1
nonwhite*|  -.0731411      .07414   -0.99   0.324  -.218446  .072164         1
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

. mfx compute, at(educ=12.56274, female=1, nonwhite=1)

Marginal effects after logit
      y  = Pr(servocc) (predict)
         =  .13425158
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
    educ |  -.0192762      .00771   -2.50   0.012  -.034393 -.004159   12.5627
  female*|   .0759481      .03483    2.18   0.029   .007681  .144215         1
nonwhite*|  -.0563873      .05558   -1.01   0.310  -.165329  .052555         1
------------------------------------------------------------------------------
```

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

. mfx compute, at(educ=15, female=1, nonwhite=1)

Marginal effects after logit
      y  = Pr(servocc) (predict)
         =  .09380016
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
    educ |  -.0140974      .00549   -2.57   0.010  -.024857 -.003338        15
  female*|   .0541132      .02723    1.99   0.047   .000743  .107483         1
nonwhite*|  -.0420635      .04077   -1.03   0.302  -.121963  .037836         1
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Note that the marginal effects reports the effect on the *probability of working in a service occupation*, while the odds ratio coefficients report the effects on the *odds of working in a service occupation*.

V. Maximum Likelihood Estimation

We cannot use OLS to estimate **probit** and **logit** equations because they are not linear in their parameters. Coefficients in these models are estimated using the *maximum likelihood estimation* (MLE).

MLE chooses coefficient estimates that maximize the likelihood of the sample data set being observed. In large samples, the coefficients obtained using MLE are unbiased, efficient, normally distributed.

A brief description of maximum likelihood estimation.

- Assume that Y=1 N_1 times and that Y=0 N_2 times (N_1+N_2=N). Also, assume that the data are ordered so that the Y=1 observations come first.
- The goal of MLE is to maximize the *joint* probability of the sample data set being observed.
- That is, MLE maximizes L = Prob(Y_1,…,Y_N). This is the same as maximizing L=Prob(Y_1) x Prob(Y_2) x … x Prob (Y_N).
- Given the way we ordered the data above, this is the same as maximizing:
  - $L = P_1 \cdots P_{N_1}(1 - P_{N_1+1}) \cdots (1 - P_N)$
- Taking logs of both sides yields the objective of maximizing the following:

    - $\log L = \sum_{i=1}^{N_1} \log P_i + \sum_{i=N_1+1}^{N} \log(1 - P_i)$

- You then plug in the appropriate equation for the P_i's, depending on whether you are estimating a logit or probit model. MLE finds $\overset{\wedge}{\beta}_0$, $\overset{\wedge}{\beta}_1$, … $\overset{\wedge}{\beta}_k$ to maximize log L above.

VI: Inference in Probit and Logit models

The log likelihood above is analogous to the SSR in OLS models. If we want to test the null hypothesis that holding race constant, education and gender do not explain working in the service occupation. Or formally,

H_0: $\beta_{educ} = 0$ and $\beta_{female} = 0$
H_1: $\beta_{educ} \neq 0$ **and/or** $\beta_{female} \neq 0$

As with the F-test, the null hypothesis imposes two restrictions so that the restricted model is now

Restricted Model: $P(Y=1|X) = \Phi(\beta_0 + \beta_3 Non\text{-}white)$ while the

Unrestricted Model: $P(Y=1|X) = \Phi(\beta_0 + \beta_1 Educ + \beta_2 Female + \beta_3 Non\text{-}white)$

Instead of an F-test, we run a log-likelihood ratio test (LR-test)

LR-stat $= -2(LL_{restricted} - LL_{unrestricted}) \sim \lambda_q$ where q is the number of restrictions

We can implement this test in STATA by typing the following

```
. probit servocc  educ female nonwhite
. estimates store A
. probit servocc  educ female nonwhite
Iteration 0:   log likelihood = -213.66328
Iteration 1:   log likelihood = -200.05342
Iteration 2:   log likelihood = -199.83821
Iteration 3:   log likelihood = -199.83803
Iteration 4:   log likelihood = -199.83803
Probit regression                                 Number of obs   =        526
                                                  LR chi2(3)      =      27.65
                                                  Prob > chi2     =     0.0000
Log likelihood = -199.83803                       Pseudo R2       =     0.0647
------------------------------------------------------------------------------
     servocc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  -.0945167   .0257061    -3.68   0.000    -.1448997   -.0441338
      female |   .4976532   .1432958     3.47   0.001     .2167985    .7785078
    nonwhite |  -.2379628   .2480437    -0.96   0.337    -.7241196    .2481939
       _cons |  -.1807477   .3283512    -0.55   0.582    -.8243043    .4628089
```

```
. probit servocc  nonwhite
. estimates store B
```

. probit servocc nonwhite

```
Iteration 0:   log likelihood = -213.66328
Iteration 1:   log likelihood = -213.43329
Iteration 2:   log likelihood = -213.43288
Iteration 3:   log likelihood = -213.43288
```

```
Probit regression                                 Number of obs   =        526
                                                  LR chi2(1)      =       0.46
                                                  Prob > chi2     =     0.4973
Log likelihood = -213.43288                       Pseudo R2       =     0.0011

------------------------------------------------------------------------------
     servocc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nonwhite |  -.1584198   .2367734    -0.67   0.503    -.6224873    .3056476
       _cons |  -1.062221   .0712249   -14.91   0.000    -1.201819   -.9226222
------------------------------------------------------------------------------
```

```
lrtest A B, stats

Likelihood-ratio test                                 LR chi2(2)  =     27.19
(Assumption: B nested in A)                           Prob > chi2 =    0.0000

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           B |    526   -213.6633   -213.4329      2     430.8658    439.3964
           A |    526   -213.6633    -199.838      4     407.6761    424.7373
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note
```

LR-stat $= -2(LL_{restricted} - LL_{unrestricted})$~ λ_q = -2(-213.433 - -199.838) = 2*13.595 = 27.19

We can reject the null that education and gender do not explain the likelihood of working in a service occupation.

There are other tests that we can implement but are beyond the scope of this course.

VII: Goodness of Fit in these models.

Unlike OLS estimations, Logit and Probit regressions don’t have an R-squared measure. Instead researchers have come up with a variety of different ways of estimating goodness of fit. Below are two such measures.

The interpretation of these measures is not the same (as OLS), but they can be interpreted as an approximate variance in the outcome accounted for by the factors.

The simplest measure looks at the fraction of observations that are accurately predicted by the model (if P(Y=1|X) > 0.5, then we would predict ‘success’ and vice versa).

The most common measure is called the Pseudo R-squared (and is due to Dan McFadden).

$$\rho^2 = 1 - \frac{LL(B)}{LL(0)}$$

this value tends to be smaller than R-square and values of .2 to .4 are considered highly satisfactory.