EEP/IAS 118
Andrew Dustan
Section Handout 13

1. Linear Probability Model vs. Logit (or Probit)

We have often used binary ("dummy") variables as explanatory variables in regressions. What about when we want to use binary variables as the *dependent* variable?

It's possible to use OLS:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

where *y* is the dummy variable. This is called the *linear probability model*.

Estimating the equation:

$$\hat{P}(y = 1|x) = \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$$

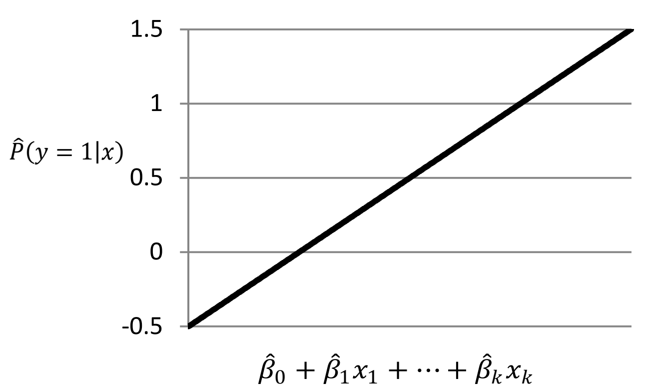
$\hat{y}$ is the predicted probability of having $y = 1$ for the given values of $x_1 \dots x_k$.

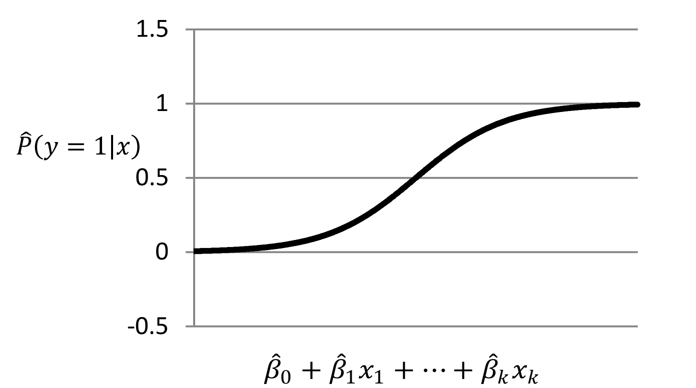
Problems with the linear probability model (LPM):

1. Heteroskedasticity: can be fixed by using the "robust" option in Stata. Not a big deal.
2. Possible to get $\hat{y} < 0$ or $\hat{y} > 1$. This makes no sense—you can't have a probability below 0 or above 1. This is a fundamental problem with the LPM that we can't patch up.

Solution: Use the *logit* or *probit* model. These models are specifically made for binary dependent variables and always result in $0 < \hat{y} < 1$. Let's leave the technicalities aside and look at a graph of a case where LPM goes wrong and the logit works:

Linear Probability Model	Logit (probit looks similar)





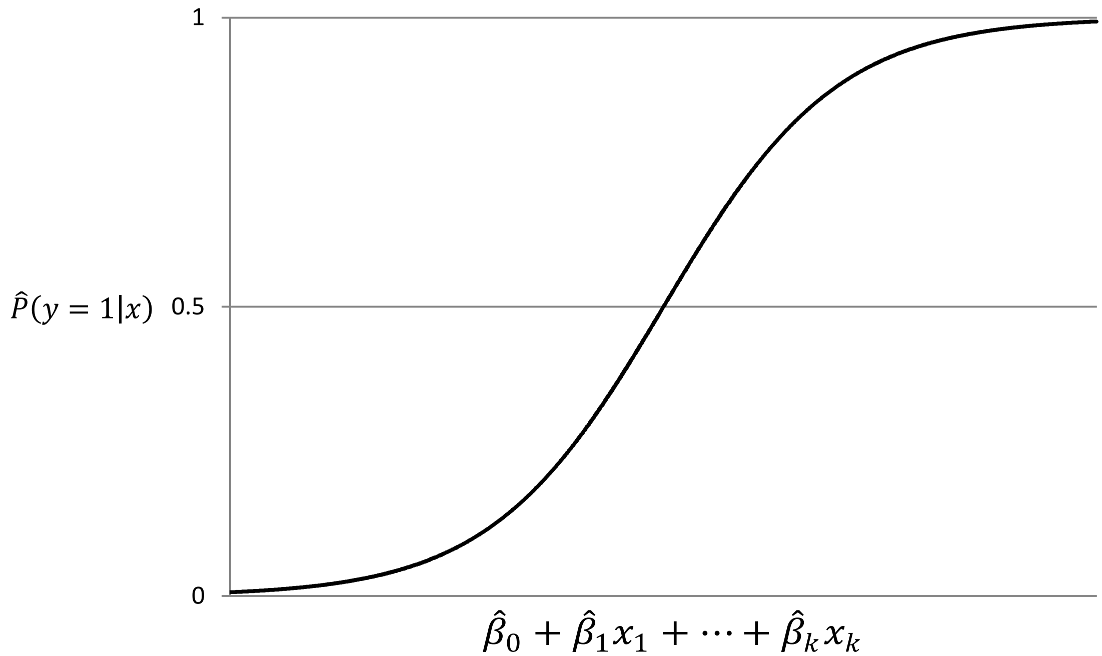
This is the main feature of a logit/probit that distinguishes it from the LPM – predicted probability of $y = 1$ is never below 0 or above 1, and the shape is always like the one on the right rather than a straight line.

2. Marginal Effects for Logit (or Probit)

We talked about how to estimate the logit using "maximum likelihood" in lecture, which is fairly complicated—much more complicated than OLS. Moreover, the results from the estimation are not easy to interpret.

What we *want* are results that look like those from OLS or the LPM: the marginal effect of changing x on $\hat{P}$, the probability of getting $y = 1$.

"Problem": the marginal effect is different depending on what the x values are. Look again at the graph:



How much does $\hat{P}$ change as we increase $\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$ (i.e. how big are marginal effects) when:

$\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$ is very low? ____a little__________

$\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$ is neither high nor low? _____a lot__________

$\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$ is very high? _____a little__________

We compromise by finding the marginal effect for the "average" person/whatever in the data, i.e. the marginal effect when $x_1 = \bar{x}_1, \ldots, x_k = \bar{x}_k$. This is what the Stata command "mfx" does.

Example: Probability of a male adult being arrested, as a function of income (in $100) and minority status:

```
. logit arrest minority inc86

Logistic regression                               Number of obs   =       2725
                                                  LR chi2(2)      =     152.22
                                                  Prob > chi2     =     0.0000
Log likelihood = -1532.0747                       Pseudo R2       =     0.0473

------------------------------------------------------------------------------
      arrest |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    minority |   .5853512   .0886866     6.60   0.000     .4115286    .7591738
       inc86 |  -.0074475   .0008404    -8.86   0.000    -.0090947   -.0058003
       _cons |  -.8499352    .069239   -12.28   0.000    -.9856411   -.7142294
------------------------------------------------------------------------------
```

The signs of these coefficients tell us something: minorities are more likely to be arrested, and higher income lowers the probability of being arrested. How big are these effects? Run "mfx" to find out:

```
. mfx

Marginal effects after logit
      y  = Pr(arrest) (predict)
         =  .26160966
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
minority*|   .1165188        .018    6.47   0.000   .081238    .1518   .378716
   inc86 |  -.0014386      .00016   -9.17   0.000  -.001746 -.001131    54.967
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Practice:

1. For males with the average level of income in this sample ($5497 in 1986 dollars), how much more likely are minorities to be arrested? (Notice that for dummy variables, Stata calculates the change from going from 0 to 1.)

11.7%

2. For males with the average level of income in this sample, how does a $1000 increase in income affect the predicted probability of being arrested?

$-.0014 \times 10 = -0.014 = -1.4\%$, so 1.4% less likely to be arrested.

3. Tests for Parameters

For linear regression, we used the t-test for the significance of one parameter and the F-test for the significance of multiple parameters. There are similar tests in the logit/probit models.

One parameter: z-test

Do this just the same way as a t-test with infinite degrees of freedom. You can read it off of the *logit*/*probit* estimation results, or the *mfx* results. The formula for testing $H_0: \hat{\beta}_j = 0$ is, just like for a t-test:

$$z = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

Practice:
Can we reject the null hypothesis of $\beta_{inc86} = 0$ at the 1% significance level?:

Yes, since $z = -9.17$, so we comfortably reject the null hypothesis.

Multiple parameters: likelihood ratio test
With the F-test, we estimated the restricted and unrestricted models, and then compared their goodness of fit (R^2). We don't have an R^2 for logit or probit, so we compare the "log likelihood" instead. The log likelihood doesn't have much meaning for us, except for this test. The closer the log likelihood gets to zero (it's always negative), the better the model fits.

To perform the likelihood ratio test, estimate the restricted (fewer variables) and unrestricted (more variables) models and then construct the test statistic:

$$LR = 2(\log\mathcal{L}_{\mathrm{U}} - \log\mathcal{L}_{\mathrm{R}})$$

where $\mathcal{L}_U$ is the likelihood from the unrestricted model and $\mathcal{L}_R$ is from the restricted model. The test statistic is distributed $\chi^2(q)$ where q is the number of restrictions, just like in the F-test. If LR is higher than the critical value, we reject the null hypothesis. This is exactly like the F-test but using the χ^2 table instead of the F table.

Practice:
We can add two variables to the arrest model: total time spent in prison in the past, and average sentence length from previous sentences (if any):

```
. logit arrest minority inc86 tottime avgsen

Logistic regression                               Number of obs   =       2725
                                                  LR chi2(4)      =     154.89
                                                  Prob > chi2     =     0.0000
Log likelihood = -1530.7407                       Pseudo R2       =     0.0482

------------------------------------------------------------------------------
      arrest |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    minority |   .5956365   .0891583     6.68   0.000      .4208894    .7703836
       inc86 |  -.0075452   .0008458    -8.92   0.000     -.0092029   -.0058875
     tottime |   -.035892     .02659    -1.35   0.177     -.0880074    .0162233
      avgsen |   .0332144   .0334359     0.99   0.321     -.0323187    .0987474
       _cons |  -.8407443   .0696835   -12.07   0.000     -.9773215   -.7041672
------------------------------------------------------------------------------
```

Do these new variables help to predict arrest, after controlling for minority status and income?

Step:	
1: Write hypotheses	$H_0: \beta_{tottime} = \beta_{avgsen} = 0$ $H_1: Not\ H_0$
2: Compute LR	$LR = 2[-1530.74 - (-1532.07)] = 2.66 \sim \chi^2(2)$
3: Get critical value	$c_{.05} = 5.99$
4: Reject/fail to reject	$2.66 < 5.99$ so fail to reject the null hypothesis
5: Conclude	We have no evidence that time spent in prison and average sentence length from previous sentences help to predict future imprisonment, after controlling for minority status and income.