5. Nonlinear Models and Limited Dependent Variables
Part 2: Limited Dependent Variables

LPO.7870: Research Design and Data Analysis II

Sean P. Corcoran

Last time

Nonlinear regression models: when the change in y for a unit change in x (i.e., the slope) changes, depending on the value of x or other regressors.

- Nonlinear functions of x (e.g., quadratic or cubic).
- Logarithms: y, x, or both can be transformed using a natural logarithm. This converts changes into *percentage* changes.
- Interaction effects: when the slope on one variable (e.g., x_1) depends on the level of another variable (e.g., x_2). The "interacted" variables may be binary or continuous.

Interaction effects are useful for testing for differential effects of a policy, treatment, or intervention by group.

Tonight

Limited dependent variables: when y is not a continuous variable, but takes on a limited range or set of values:

- Binary outcomes (0-1)
- Categorical outcomes

These outcome variables do not fit the standard regression model and may require some different tools.

Tonight's sample datasets

We will refer to two datasets tonight (on Github):

1. `nels.dta`: 8th grade students in the NELS.
2. `hmda.dta`: Boston Fed Home Mortgage Disclosure Act (HMDA) data. 2,380 mortgage applicants in 1990, data collected by the Boston Fed. For single-family residences only, and only Black and White applicants. For more information, see Munnell et al. (1996).

The linear probability model

Limited dependent variables

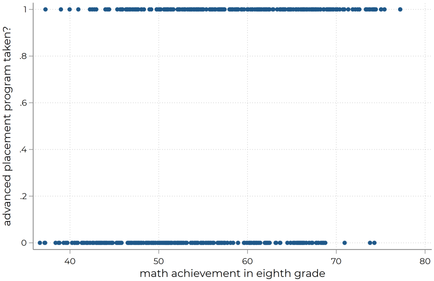
A **limited dependent variable** is when y is not continuous but takes on a limited range or set of values. Examples:

- Graduated from high school (0-1)
- Enrolled in a four-year college (0-1)
- Teacher turnover (0-1)
- Denied a mortgage loan (0-1)
- Attended a public, private, charter, or homeschool (categorical)
- Commuted to school by car, public transport, or by foot (categorical)

The last two types are sometimes called "qualitative" dependent variables.

Predicting AP course-taking

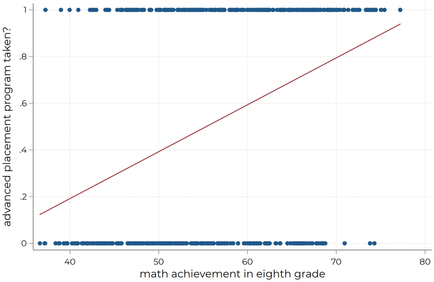
For example, consider the relationship in the NELS between taking an AP course in high school and one's 8th grade math score:



What is the "best fit" line for these data?

Predicting AP course-taking

Stata's best fit line (`lfit`) for these variables:



How should this be interpreted?

Predicting AP course-taking

The simple linear regression is:

$$approg = \beta_0 + \beta_1 achmat08 + u$$

```
. reg approg achmat08

      Source |       SS           df       MS      Number of obs   =       493
-------------+----------------------------------   F(1, 491)       =     80.23
       Model |  17.2663058         1  17.2663058   Prob > F        =    0.0000
    Residual |  105.666757       491  .215207244   R-squared       =    0.1405
-------------+----------------------------------   Adj R-squared   =    0.1387
       Total |  122.933063       492  .249863949   Root MSE        =     .4639

------------------------------------------------------------------------------
      approg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    achmat08 |   .0200978   .0022438     8.96   0.000     .0156892    .0245064
       _cons |  -.6122942   .1287168    -4.76   0.000    -.8651978   -.3593906
------------------------------------------------------------------------------
```

The slope β_1 is the change in y for a 1-unit change in x, but y is either 0 or 1 here. We can think of the slope as how much closer (or further away, if negative) we get to $y = 1$ when x changes by one unit.

Linear probability models

The above regression is called a **linear probability model** (LPM)—i.e., when using OLS with a 0-1 dependent variable.

We can interpret β_1 as the change in the *probability* that $y = 1$ when x changes by one unit. This has many useful applications!

- How is GPA related to the probability of graduating high school?
- To what extent does a higher SAT score predict enrollment in a four-year college?
- Is a teacher more likely to exit a school with concentrated poverty?
- Are Black applicants more likely to be denied a home mortgage, all else equal?

Linear probability models

Interpretation:

```
. reg approg achmat08

      Source |       SS           df       MS      Number of obs   =       493
-------------+----------------------------------   F(1, 491)       =     80.23
       Model |  17.2663058         1  17.2663058   Prob > F        =    0.0000
    Residual |  105.666757       491  .215207244   R-squared       =    0.1405
-------------+----------------------------------   Adj R-squared   =    0.1387
       Total |  122.933063       492  .249863949   Root MSE        =     .4639

------------------------------------------------------------------------------
      approg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    achmat08 |   .0200978   .0022438     8.96   0.000     .0156892    .0245064
       _cons |  -.6122942   .1287168    -4.76   0.000    -.8651978   -.3593906
------------------------------------------------------------------------------
```

A one-point increase in the 8th grade math score is associated with a 2 **percentage point increase** in the probability of taking an AP course in high school.

Linear probability models

Attenzione! this is generally not the same as "2%" increase!! This statement implies a change from some baseline.

```
. summ approg achmat08

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      approg |        493     .525355    .4998639          0          1
    achmat08 |        500    56.59102    9.339608      36.61       77.2

. di 0.75*(1.02)
.765

. di 0.765 - 0.75
.015
```

If Student A had a 75% probability of taking AP, a 2% increase would move her to 76.5%. This is a 1.5 percentage point change. The slope in a LPM is a percentage point change. Be careful in how you interpret!

Linear probability models

An aside: is a 2 percentage point increase in AP course taking a *practically significant* one?

First consider a meaningful change in 8th grade math achievement: the standard deviation of *achmat08* is **9.34**. If a one-point increase in *achmat08* is associated with a 2 percentage point increase in AP, a 9.34 point increase is associated with a **18.7** percentage point increase in AP.

Is that big? Well, 52.5% of students overall take AP. An 18.7 percentage point increase on this baseline is pretty sizable (it's about a 36% increase).

Takeaway: 8th grade math achievement is a strong predictor of AP course-taking!

Linear probability models

Predicted values in the LPM are just the predicted *probability* that $y = 1$ given a value of x:

$$\widehat{approg} = \widehat{\beta_0} + \widehat{\beta_1} achmat08$$

For example, for a student who scored a 70 in 8th grade math:

```
. ** predicted probability of AP given a math score of 70
. margins, at(achmat08=70)

Adjusted predictions                          Number of obs     =        493
Model VCE    : OLS

Expression   : Linear prediction, predict()
at           : achmat08        =          70

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .7945517   .0366026    21.71   0.000     .7226346    .8664689
------------------------------------------------------------------------------
```

Predicting AP course-taking

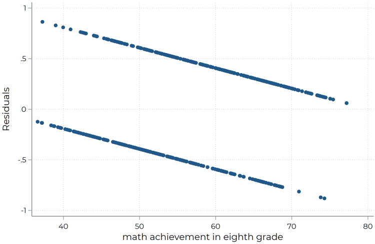
Since the LPM is just an OLS regression model, it can be extended to multiple explanatory variables:

```
. reg approg achmat08 i.gender i.urban ses

      Source |       SS           df       MS      Number of obs   =       493
-------------+----------------------------------   F(5, 487)       =     21.79
       Model |  22.4749882         5  4.49499765   Prob > F        =    0.0000
    Residual |  100.458075       487  .206279414   R-squared       =    0.1828
-------------+----------------------------------   Adj R-squared   =    0.1744
       Total |  122.933063       492  .249863949   Root MSE        =    .45418

------------------------------------------------------------------------------
      approg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    achmat08 |   .0171135   .0023525     7.27   0.000     .0124913    .0217358
             |
      gender |
     female  |   .0016509   .0416118     0.04   0.968    -.0801099    .0834117
             |
       urban |
   suburban  |  -.0348284   .0519669    -0.67   0.503    -.1369355    .0672787
      rural  |  -.2050885   .0565681    -3.63   0.000    -.3162361   -.0939409
             |
         ses |   .0063168   .0032086     1.97   0.050     .0000124    .0126212
       _cons |   -.479396   .1459171    -3.29   0.001    -.7661007   -.1926912
------------------------------------------------------------------------------
```

How should we interpret these coefficients?

Problems with LPM (1): heteroskedasticity

The LPM has some technical issues that can present problems. For one, the errors are heteroskedastic. Recall that residuals are the actual y minus the predicted ($\hat{u} = y - \hat{y}$), and in the LPM the actual y is only 0 or 1:

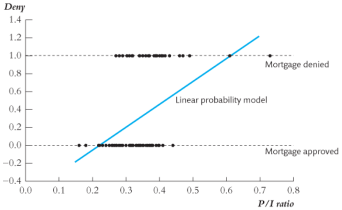


Not a big deal if we adjust our standard errors appropriately (`robust`).

Problems with LPM (2): nonsensical predictions

More seriously, the LPM can generate some nonsensical predictions: predicted probabilities that are < 0 or > 1:

Figure 11.1 Scatterplot of Mortgage Application Denial and the Payment-to-Income Ratio



Mortgage applicants with a high ratio of debt payments to income (*P/I ratio*) are more likely to have their application denied (*deny* = 1 if denied; *deny* = 0 if approved). The linear probability model uses a straight line to model the probability of denial, conditional on the *P/I ratio*.

Source: Stock & Watson chapter 11.

Problems with LPM (2): nonsensical predictions

Consider our earlier LPM for AP course-taking. Get predicted probabilities at *achmat08* of 25 and 85:

```
. margins, at(achmat08=25)

Adjusted predictions                            Number of obs     =        493
Model VCE     : OLS

Expression    : Linear prediction, predict()
at            : achmat08        =           25

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |  -.1098492   .0739294    -1.49   0.138    -.2551062    .0354078
------------------------------------------------------------------------------

. margins, at(achmat08=85)

Adjusted predictions                            Number of obs     =        493
Model VCE     : OLS

Expression    : Linear prediction, predict()
at            : achmat08        =           85

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   1.096019   .0670486    16.35   0.000     .9642811    1.227756
------------------------------------------------------------------------------
```

Problems with LPM

In the next section we will see alternatives to the LPM, but in practice researchers have found that it works well in many applications. You will often see analysts use the LPM despite its technical concerns.

Sample paper: Ladd (2011)

Ladd (2011): Teachers' Perceptions of their Working Conditions: How Predictive of Planned and Actual Teacher Movement?

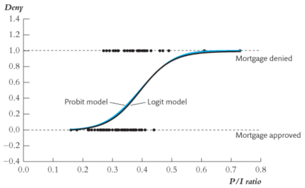
- Uses North Carolina administrative data on teachers combined with a 2006 statewide survey given to all teachers in the state.
- Relates teachers' perceptions of working conditions to *intended* and *actual* movement from their schools, controlling for other school characteristics.
- Teacher mobility is a binary variable, so Ladd uses a LPM
- Note working conditions variables are z-scores (mean 0 sd 1)
- See results in Tables 3, 4, 5. Full paper is on Github.

Logit and probit models

Logit and probit models

The fact that the probability $y = 1$ is bounded between 0 and 1 suggests we need a nonlinear regression model. Logit (logistic) and probit are two functional forms that meet this requirement:

Figure 11.3 Probit and Logit Models of the Probability of Denial Given *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.

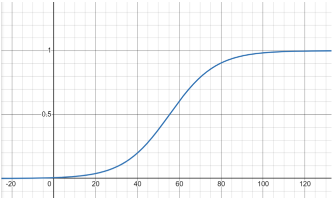
Source: Stock & Watson chapter 11.

Logistic regression

The logistic model expresses the probability that $y = 1$ as function of x:

$$Pr(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

For example, if $\beta_0 = -5$ and $\beta_1 = 0.091$ (these values come up later):



Source: `https://www.desmos.com/calculator`

Logistic regression

The idea is to find parameters β_0 and β_1 that provide the "best fit" logistic curve to the given data.

The problem is that this function does not fit with our standard regression model (linear in coefficients). We can't estimate the parameters using ordinary least squares!

Maximum likelihood

Solution: logit parameters can be estimated using **maximum likelihood** (ML). The details are beyond the scope of this class—but in short:

- For observations where $y = 1$ the probability is $Pr(y = 1|x)$, which was given above, assuming the logistic function.
- For observations where $y = 0$ the probability is $1 - Pr(y = 1|x)$.
- The joint probability of observing the data you observed is the product of these probabilities for all observations. (Assumes the observations are independent).
- ML finds the β_0 and β_1 most consistent with the observed data.

Don't worry, Stata will do this for you!

Predicting AP course-taking, revisited

Let's try estimating the AP course-taking model assuming a logit function (use the `logit` command in Stata instead of `regress`):

```
. logit approg achmat08

Iteration 0:   log likelihood = -341.08741
Iteration 1:   log likelihood =  -304.3896
Iteration 2:   log likelihood = -304.36471
Iteration 3:   log likelihood = -304.36471

Logistic regression                             Number of obs     =        493
                                                LR chi2(1)        =      73.45
                                                Prob > chi2       =     0.0000
Log likelihood = -304.36471                     Pseudo R2         =     0.1077

------------------------------------------------------------------------------
      approg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    achmat08 |   .0908892   .0115717     7.85   0.000      .068209    .1135693
       _cons |  -5.021598   .6566153    -7.65   0.000     -6.30854   -3.734655
------------------------------------------------------------------------------
```

Notice these numbers are the ones I used earlier (-5, 0.091).

Predicting AP course-taking, revisited

Compare these coefficients to those in the LPM—they are very different!

```
. reg approg achmat08

      Source |       SS           df       MS      Number of obs   =       493
-------------+----------------------------------   F(1, 491)       =     80.23
       Model |  17.2663058         1  17.2663058   Prob > F        =    0.0000
    Residual |  105.666757       491  .215207244   R-squared       =    0.1405
-------------+----------------------------------   Adj R-squared   =    0.1387
       Total |  122.933063       492  .249863949   Root MSE        =     .4639

------------------------------------------------------------------------------
      approg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    achmat08 |   .0200978   .0022438     8.96   0.000     .0156892    .0245064
       _cons |  -.6122942   .1287168    -4.76   0.000    -.8651978   -.3593906
------------------------------------------------------------------------------
```

This is because the coefficients are not directly comparable. In the LPM, β_1 is the change in $Pr(y=1)$ for a 1-unit change in x. The slope is constant. The logistic function is nonlinear (an "S" curve). The change in $Pr(y=1)$ for a 1-unit change in x is <u>not</u> β_1—it's more complicated.

Bottom line: don't interpret logit coefficients as slopes!

Predicting AP course-taking, revisited

So what can you do? Stata's `margins` command comes in handy:

```
. margins ,at(achmat08=56.6) dydx(achmat08)

Conditional marginal effects                    Number of obs     =        493
Model VCE    : OIM

Expression   : Pr(approg), predict()
dy/dx w.r.t. : achmat08
at           : achmat08        =        56.6

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    achmat08 |   .0226369   .0028775     7.87   0.000     .0169972    .0282767
------------------------------------------------------------------------------

. margins ,at(achmat08=65) dydx(achmat08)

Conditional marginal effects                    Number of obs     =        493
Model VCE    : OIM

Expression   : Pr(approg), predict()
dy/dx w.r.t. : achmat08
at           : achmat08        =          65

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    achmat08 |   .0187861   .0017609    10.67   0.000     .0153348    .0222374
------------------------------------------------------------------------------
```

These are pretty close to 0.02. But notice that the slope depends on the value of *achmat08*.

Using odds ratios

As it turns out, the logit function tells you something about the "odds" of y being equal to 1. Let p be $Pr(y = 1)$:

$$ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$$

$p/(1-p)$ is the **odds** that $y = 1$ and $ln(p/(1-p))$ is the **log-odds**.

Using odds ratios

The odds of an event is the ratio of the expected number of times an event *will* occur to the expected number of times an event will *not* occur.

- An odds of 10 means we expect 10 times as many occurrences as non-occurrences
- Odds ratios greater than 1 correspond to positive predictive effects of an explanatory variable.
- Odds ratios less than 1 correspond to negative predictive effects of an explanatory variable.

Using odds ratios

Stata's `logistic` command (different from `logit`) will provide odds ratios:

```
. logistic approg achmat08

Logistic regression                             Number of obs     =        493
                                                LR chi2(1)        =      73.45
                                                Prob > chi2       =     0.0000
Log likelihood = -304.36471                     Pseudo R2         =     0.1077

------------------------------------------------------------------------------
      approg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    achmat08 |   1.095148   .0126727     7.85   0.000     1.070589     1.12027
       _cons |    .006594   .0043297    -7.65   0.000     .0018207    .0238814
------------------------------------------------------------------------------
Note: _cons estimates baseline odds.
```

A 1-point increase in the 8th grade math score increases the *odds* of taking AP by 9.5%. This is easier to interpret than the logit coefficients themselves, but I still find odds ratios difficult to think about. (Some fields predominately work with odds ratios).

Probit model

The probit model expresses the probability that $y = 1$ as a different function of x:

$$Pr(y = 1|x) = \Phi(\beta_0 + \beta_1 x)$$

where Φ is the cumulative standard normal distribution:

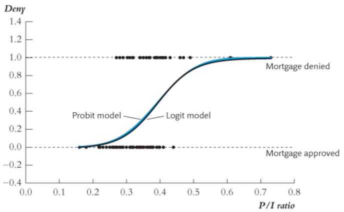
$$Pr(y = 1|x) = \frac{1}{2\pi} \int_{-\infty}^{\beta_0+\beta_1 x} \exp\left(-\frac{u^2}{2}\right) du$$

This is a more complicated-looking function than the logit, but the idea is the same, and ML is used to find β_0 and β_1.

Probit model

The fitted probit often looks very similar to the logit:

Figure 11.3 Probit and Logit Models of the Probability of Denial Given *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.

In practice it often doesn't make a big difference which choice you make.

Predicting AP course-taking, revisited

Let's try estimating the AP course-taking model assuming a probit function (use the `probit` command in Stata instead of `regress`):

```
. ** try a probit function
. probit approg achmat08

Iteration 0:   log likelihood = -341.08741
Iteration 1:   log likelihood = -304.19248
Iteration 2:   log likelihood =  -304.1737
Iteration 3:   log likelihood =  -304.1737

Probit regression                               Number of obs     =        493
                                                LR chi2(1)        =      73.83
                                                Prob > chi2       =     0.0000
Log likelihood =  -304.1737                     Pseudo R2         =     0.1082

------------------------------------------------------------------------------
      approg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    achmat08 |   .0559988   .0068266     8.20   0.000     .0426188    .0693788
       _cons |  -3.093653    .388287    -7.97   0.000    -3.854681   -2.332624
------------------------------------------------------------------------------
```

Again the coefficients look very different from LPM and logit. Do not interpret these directly as slopes!

Predicting AP course-taking, revisited

Compare the slopes calculated at the same points as earlier:

```
. margins ,at(achmat08=56.6) dydx(achmat08)

Conditional marginal effects                    Number of obs     =        493
Model VCE    : OIM

Expression   : Pr(approg), predict()
dy/dx w.r.t. : achmat08
at           : achmat08        =        56.6

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    achmat08 |   .0222761    .002713     8.21   0.000     .0169587    .0275934
------------------------------------------------------------------------------

. margins ,at(achmat08=65) dydx(achmat08)

Conditional marginal effects                    Number of obs     =        493
Model VCE    : OIM

Expression   : Pr(approg), predict()
dy/dx w.r.t. : achmat08
at           : achmat08        =          65

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    achmat08 |   .0192438   .0018248    10.55   0.000     .0156673    .0228203
------------------------------------------------------------------------------
```


Logit and probit in publications

If you see the logistic or probit model used in a publication, make sure you pay attention to what they are reporting:

- Is it an estimated logit or probit coefficient? (If so, do not interpret directly, but the sign $+/-$ and significance may be informative).
- Is it an odds ratio? (If so, see if the odds ratio is greater than or less than 1).
- Is it a *slope* calculated from the logit/probit coefficients? If so, at what values of x are the calculating the slope? Remember, the slope changes in these models, depending on the values of x.

Personally, I prefer authors calculate the slopes for me at policy-relevant values of x.

Exercise using HMDA data

Is there racial discrimination in mortgage lending?

1. What is the difference in rejections (*deny*=1) between Black and other applicants?

2. One factor that may explain differences in rejections is the borrower's debt-to-income ratio (*pi_rat*). Regress *deny* on *pi_rat* and interpret. Then, add *black*. Is the latter statistically significant?

3. Now add more controls: housing expense to income ratio (*hse_inc*), dummy variables for medium and high loan-to-house value ratios (*ltv_med*, *ltv_high*), consumer credit score (*ccred*), mortgage credit score (*mcred*), public bad credit record (*pubrec*), denied mortgage insurance (*denpmi*), and self employed (*selfemp*). How does this effect the role of Black in the regression?

Try the above regressions using LPM and logit and compare.

Next time

Experimental and quasi-experimental methods: causal inference through good research design.

- Randomized controlled trials
- Regression discontinuity designs (Lecture 7)
- Difference-in-differences designs (Lecture 8)
- Panel data methods (Lecture 9)
- And much, much, more!