EEP/IAS 118 - Introductory Applied Econometrics, Lecture 13

Gregory Lane

August 2017

Review: Impact Evaluation Methods

1. **Randomization (RCT)**
    - *Key Assumption:* If it were not for the treatment, C and T group would be statistically identical
    - *Test of Assumption:* Test “balance” of variables not affected by the treatment
2. **Diff-in-Diff**
    - *Key Assumption:* change from before to after in the comparison group is a good counterfactual for the treatment group
    - *Test of Assumption:* Check for “parallel trends” in the periods before the treatment

Review: Impact Evaluation Methods

3. **Generalized Diff-in-Diff**
   - *Key Assumption:* Annual change in control group(s) is good counterfactual for annual change in treatment group
   - *Test of Assumption:*
     1. Entry into treatment not correlated with pre-trends
     2. Absence of “Ashenfelter Dip”
     3. Robustness to other policies
4. **Regression Discontinuity**
   - *Key Assumption:* The outcome would be a continuous function of the running variable around the threshold, if it were not for the treatment
   - *Test of Assumption:* Check for absence of discontinuity in other variables around the threshold

Linear Probability Model: Intro

Previously, we have talked extensively about including dummy variables as X variables. However, sometimes our Y variable can also be a dummy (i.e. only takes values of 0 or 1)

Let's say we don't change anything and run traditional linear regression with with a dummy variable as our Y:

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

where

$$y \in \{0, 1\}$$

We can this type of analysis, a "linear probability model"

Linear Probability Model: Interpretation

What changes with a dummy variable as our Y?

- We **can't** interpret β_j as the unit change in y given a one-unit increase in x_j holding all other factors fixed
- Y either changes from $0 \to 1$ or $1 \to 0$ or doesn't change
- Instead, β_j measures the change in the **probability** of Y taking the value 1 when x_j changes by one unit holding all other factors fixed
- And, $\hat{\beta}_j$ measures the **predicted** change in the **probability** of success when x_j increases by one unit holding all other variables fixed

Linear Probability Model: Example

Let's say we wanted to study the probability of women working outside the home. We run a regression with a binary variables as the Y indicating a woman was working outside the home on several explanatory variables:

$$\widehat{inlf} = 0.586 - 0.0034 nwifeinc + 0.38 educ + 0.039 exper - 0.0060 exper^2 - 0.017 age - 0.262 kindslt6 + 0.0130 kidsge6$$

How do we interpret the coefficient on education?

Linear Probability Model: Example

Let's say we wanted to study the probability of women working outside the home. We run a regression with a binary variables as the Y indicating a woman was working outside the home on several explanatory variables:

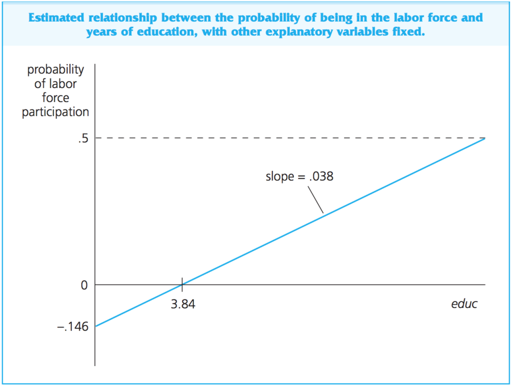
$$\begin{aligned}\widehat{inlf} &= 0.586 - 0.0034 nwifeinc + 0.038 educ + 0.039 exper \\ &\quad - 0.0060 exper^2 - 0.017 age - 0.262 kindslt6 + 0.0130 kidsge6\end{aligned}$$

How do we interpret the coefficient on education?

Another year of education increases the predicted probability of labor force participation by 0.038 holding all else constant

Linear Probability Model: Example

We fix the values of the other variables and graph the relationship:



- It is possible to get negative probabilities
- The marginal effect of an additional year of education on the probability of participation is constant (at 0.038)

Drawbacks of Linear Prob. Model

1. Predicted probabilities from regression aren't bounded between zero and one
2. There must be heteroskedasticity in the linear probability model. This violates our assumption of homoskedasticity:

$$Var(u|x) = Var(u) = \sigma^2$$

   Therefore, our standard error calculations are more difficult

Maximum Likelihood

How do we address these drawbacks? Use non-linear, maximum-likelihood (ML) technics rather than OLS

There are many different types of ML techniques depending on the specific situation:

- **Binary Y**: Logit or Probit model
- **Count Data:** Poisson Model
- **Truncated Data:** $(a \leq y \leq b)$ Tobit Model
- **Order Categorical Data:** (1-5 rating scale) Ordered Probit

We will only focus on the first case

Logit and Probit

The logit and probit functions look like this:

- Logit:

$$Pr(y_i = 1 \mid x_{1i}, x_{2i}) = \frac{e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}}}$$

- Probit:

$$Pr(y_i = 1 \mid x_{1i}, x_{2i}) = \int_{-\infty}^{\beta + \beta_1 x_{1i} + \beta_2 x_{2i}} \phi(v) dv$$

- Both output a probability for $y = 1$ given an input of x values
- **Difference:** Logit using the logistic distribution, probit uses the normal distribution
- Both estimate $\hat{\beta}_j$ by maximizing the log-likelihood function

Latent Variable Framework

How can we think about these equations? It is easiest to understand this through the latent variable framework:

- Assume there is a variable y^* that is generated by:

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

We do *not* observe y^*. Instead we only see y which is simply:

$$\begin{cases} y = 1 \text{ if } y^* \geq 0 \\ y = 0 \text{ if } y^* < 0 \end{cases}$$

I.e. think about a purchase of a good based on demand. We only observe a purchase ($y = 1$) if demand (y^* is greater than the price $y^* \geq p$)

Latent Variable Framework

What the Logit and Probit models are doing is making and assumption on the error term u in the latent variable regression.

- **Logit:** u has logistic distribution
- **Probit:** u has normal distribution

Start with assuming u has a normal distribution:

$$\begin{aligned} Prob(y = 1) &= Prob(y^* = \beta_0 + X\beta + u > 0) \\ &= Prob(-u < \beta_0 + X\beta) \\ &= \Phi(\beta_0 + X\beta) \\ &= Probit \end{aligned}$$

Where $\Phi(\cdot)$ is the CDF for the normal distribution

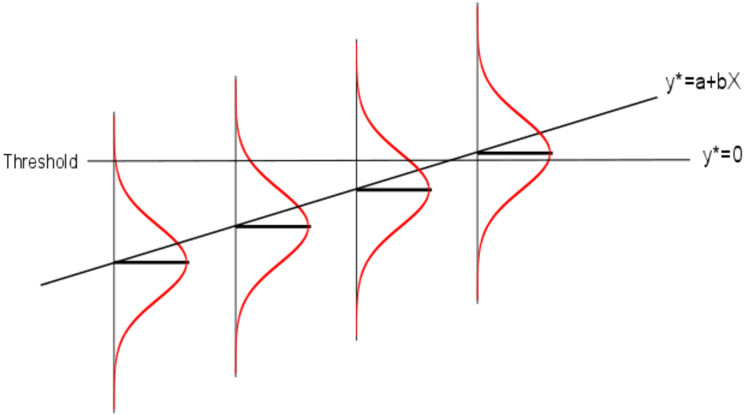
Latent Variable Framework

Likewise, if u follows an extreme value distribution:

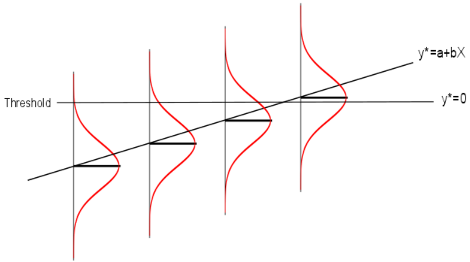
$$\begin{aligned}
Prob(y = 1) &= Prob(y^* = \beta_0 + X\beta + u > 0) \\
&= Prob(-\mu < \beta_0 + X\beta) \\
&= \frac{e^{\beta_0 + X\beta}}{1 + e^{\beta_0 + X\beta}}
\end{aligned}$$

Latent Variable Framework

We can visualize what this model looks like:



Latent Variable Framework



What can this picture tell us:

- The change in probability that $y = 1$ is *not* constant
- β may not be actually what we want to know

Logit and Probit

This is complicated. Here are the main take-aways you should know:

- Logit and Probit functions are bounded between 0 and 1, so our predicted probabilities always make sense
- The *marginal effect* of x on $P(y = 1)$ depends on the values of x_j. $\Rightarrow$ there is **NOT** a constant marginal effect
- When asked for the marginal effect of x on $P(y = 1)$, we typically report their values at the mean of the Xes
- Stata does **NOT** automatically report the marginal effects. Instead, you must specify the mfx command after running a logit or probit model

Logit and Probit: Example

Examine the probability that the favored actually team wins in a game of basketball

$$favwin_i = \hat{\beta}_0 + \hat{\beta}_1 spread_i + \hat{\beta}_2 favhome_i + \hat{\beta}_3 fav25_i + \hat{\beta}_4 und25_i + \hat{u}_i$$

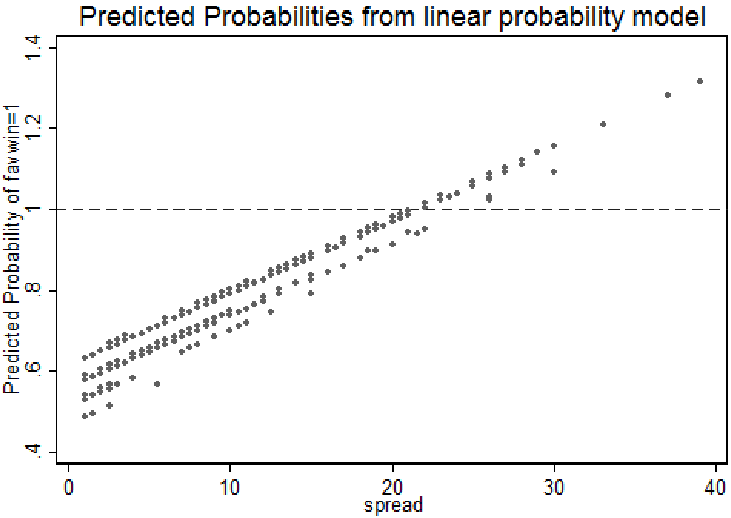
- Estimate first with a linear probability model

Linear Probability Model

```
. reg favwin spread favhome fav25 und25, robust
Linear regression                                      Number of obs =     553
                                                       F(  4,   548) =   26.20
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1160
                                                       Root MSE      =  .40158
------------------------------------------------------------------------------
        |               Robust
 favwin |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 spread |   .0177628   .0020565     8.64   0.000     .0137233    .0218023
favhome |   .0543528   .0409228     1.33   0.185    -.026032     .1347376
  fav25 |   .0109819   .0391002     0.28   0.779    -.0658228    .0877865
  und25 |  -.101104    .0895298    -1.13   0.259    -.2769676    .0747596
  _cons |   .5588152   .0404217    13.82   0.000     .4794148    .6382156
------------------------------------------------------------------------------
```

- A 1 point increase in the Vegas point spread is estimated to increase the predicted probability of winning by 1.78 percentage points, holding the other regressors constant

Predicted Win Probabilities: Linear



Probit

```
probit favwin spread favhome fav25 und25

Iteration 3:   log likelihood = -262.64177
Iteration 4:   log likelihood = -262.64177

Probit regression                                 Number of obs   =        553
                                                  LR chi2(4)      =      80.22
                                                  Prob > chi2     =     0.0000
Log likelihood = -262.64177                       Pseudo R2       =     0.1325

------------------------------------------------------------------------------
 favwin |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 spread |   .0878845   .0129491     6.79   0.000     .0625047    .1132642
favhome |   .1485753   .1370571     1.08   0.278    -.1200517    .4172024
  fav25 |    .003068     .15869     0.02   0.985    -.3079587    .3140946
  und25 |  -.2198082   .2505842    -0.88   0.380    -.7109443    .2713278
  _cons |  -.0551801    .128763    -0.43   0.668    -.3075509    .1971907
------------------------------------------------------------------------------
```

DON'T Interpret these!

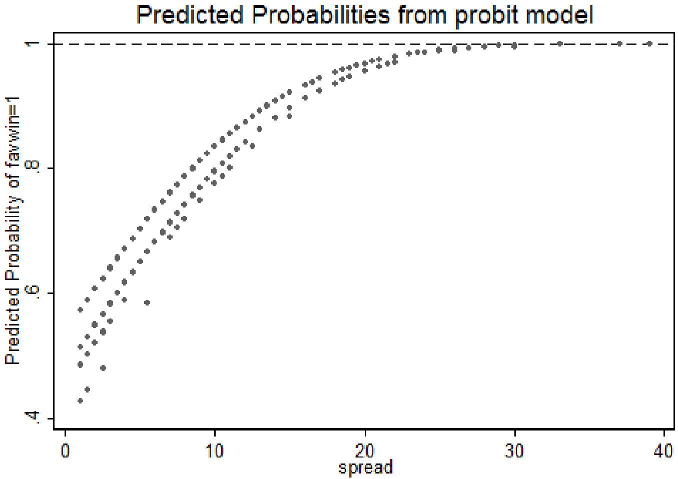
Probit: Marginal Effects

```
mfx

Marginal effects after probit
y  = Pr(favwin) (predict)
=  .80994736
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
  spread |   .0238529      .00313    7.62   0.000   .017719  .029987   9.61302
 favhome*|   .0412478      .03896    1.06   0.290  -.035104  .117599   .678119
   fav25*|   .0008322      .04302    0.02   0.985  -.083481  .085145   .264014
   und25*|  -.0645737      .07918   -0.82   0.415  -.219761  .090614   .061483
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

We can interpret as before: A 1 point increase in the Vegas point spread is estimated to increase the probability of winning by 2.39 percentage points, holding all else constant

Predicted Win Prob: Probit



Likelihood Ratio Test

The likelihood ratio test (or ChiSquared test) is *an F-test for the logit and probit models*

- Just like an F-test, an LR compares the fit of two nested regressions, an unrestricted model and a restricted model.
- The test statistic is:

$$LR = 2(LL_{UR} - LL_R)$$

  Where LL_{UR} is the "log-likelihood" from the unrestricted model, and LL_R is the log-likelihood from the restricted model
- It is distributed Chi-squared, so our critical values will be found in the chi-squared table

Likelihood Ratio Test: Example

we wanted to test the hypothesis that being either in the top 25 teams as the favored team or as the underdog doesn't affect the probability of the favored team winning. To test this hypothesis, we'd like to compare these two models:

Unrestricted model: $Pr(favwin = 1|spread...) =$

$$\Phi(\hat{\beta}_0 + \hat{\beta}_1 spread_i + \hat{\beta}_2 favhome_i + \hat{\beta}_3 fav25_i + \hat{\beta}_4 und25_i)$$

Restricted model: $Pr(favwin = 1|spread...) =$

$$\Phi(\hat{\beta}_0 + \hat{\beta}_1 spread_i + \hat{\beta}_2 favhome_i)$$

Likelihood Ratio Test: Example

Run the restricted model:

```
. probit favwin spread favhome
Iteration 0:   log likelihood = -302.74988
Iteration 1:   log likelihood = -264.51089
Iteration 2:   log likelihood = -263.07028
Iteration 3:   log likelihood = -263.06924
Iteration 4:   log likelihood = -263.06924
Probit regression                                 Number of obs   =        553
                                                  LR chi2(2)      =      79.36
                                                  Prob > chi2     =     0.0000
Log likelihood = -263.06924                       Pseudo R2       =     0.1311

------------------------------------------------------------------------------
 favwin |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
 spread |   .0900149   .0124116     7.25   0.000     .0656885    .1143413
favhome |   .1311407   .1318301     0.99   0.320    -.1272415    .3895229
  _cons |   -.075713    .122677    -0.62   0.537    -.3161554    .1647295
------------------------------------------------------------------------------
```

Log likelihood = -263.069

Likelihood Ratio Test: Example

After doing the same for the unrestricted model, we get our test statistic:

$$LR = 2(LL_{UR} - LL_R) = 2(-262.64177 - (-263.06924)) = 0.855$$

Let's say we want to test at the 10% level: $\alpha = 0.10$

- Looking at the table for two degrees of freedom ($q = 2$)
- Critical stat = 4.61

At the 10% significance level, we Fail to Reject the null hypothesis because $LR < c$.