Qualitative Dependent Variables
Housekeeping

• Paper is due the last day of class.
• Presentation signup today
How do we test for heteroskedasticity?

• After your regression, plot the residuals against the fitted values and the values of each x. (see rvfplot)

• You can also run Cameron and Trivedi’s decomposition test: estat imtest
  – Or a Breusch-Pagan/Cook-Weisberg test: estat hettest x2 x3 x4 ...., mtest

• But a large sample size and might reject the null of homoskedasticity even if heteroskedasticity is very slight.
Outline

• Nature of qualitative dependent variables
• Linear Probability Model
• Logit and Probit
• Estimating Logit and Probit
• Interpreting Logit and Probit
• Model Performance
• Further Topics
Dichotomous Dependent Variables

• Variables that have two categories
• Used to indicate whether an event has occurred or not or to indicate that some characteristic is present (basically a dummy dependent variable)
• Examples:
  – Is this country a democracy?
  – Is there a bilateral investment treaty in force for this dyad-year?
  – Was their a terrorist attack in this country-year??
  – Was their a war in this dyad-year?
Other types of qualitative DV’s

• Can have more than two choices
• Trichotomous
  – Example vote choice in 1992
  – Clinton, Bush, Perot
• Polychotomous
  – Multiple categories
  – Likert scales-5 categories
• Today we will focus on models for dichotomous Dependent Variables
  – (only take on values of 0 or 1)
Basic Difference

• Quantitative DV
  – Objective is to estimate the expected mean of $Y$ given the values of the regressors, or $X$ variables

• Qualitative DV
  – Objective is to find the probability of something happening, such as voting for the Republican candidate, given the values of the regressors, or $X$ variables
How to Model Dichotomous DVs

• $y_i = x_i b + u_i$, where $y_i = \{0, 1\}$
• we need a probability model that describes $y_i$
• Start with Linear Probability Model
  – Fits a regular OLS regression to a binary dependent variable
  – $E(Y) = Xb$
Linear Probability Model

• What is the expected value of Y in the binary case?
  – \( E(Y) = 1[Pr(Y=1)] + 0[Pr(Y=0)] \)
  – \( =Pr(Y=1) \)
    • This is the mean of a Bernoulli distribution

• We can then model: Run linear OLS on X
  – \( Pr(Y_i=1) = X_i \beta \)
    • Model will estimate properly
    • Interpret coefficients as the change in \( Pr(Y=1) \) is associated with a one unit change in X
Would look like this
Problems with LPM: Non-normality of the distribution

- OLS assumed normality of the errors for inference
- Not tenable for LPM because \( u_i \) can only take 2 values, and follows the Bernoulli distribution
  - \( 1-X_ib \) and \( -X_ib \)
- If errors not normally distributed, can not do hypothesis testing
- Not as critical because OLS remains unbiased, and as \( n \) increases to infinity, OLS estimators tend to be normally distributed
Problems with LPM: Heteroskedasticity

- Error variance is heteroskedastic, by construction
  - $\text{Var}(u_i) = P_i(1 - P_i)$
    - But $P_i = X_i B$, thus is a function of the mean
Problems with LPM: Non-fullfillment of bounds and $R^2$

• Our values of Y can only assume 0 or 1

• The LPM predictions are not bounded, so we might have probabilities less than 0 or greater than one

• Also, $R^2$ is meaningless since no LMP will fit a scatter plot well
Problems with LPM-Linearity assumption

• Even if we solve heteroskedasticity problem, we’re assuming that $b$ has a linear effect, which is unrealistic.

• Expect at ends $X$ will have less of an impact
  – There will be some cutpoint where effect is steeper
What we need

• Need Model with 2 features
  – Bounded between 0 and 1

  – Non-linear relationship between \( p_i \) and \( x_i \)
    • “one which approaches zero at slower and slower rates as \( x_i \) gets small and one which approaches 1 at slower and slower rates as \( x_i \) gets large”

    • S shaped curve (cumulative distribution function is one)
The Link Functions: Probit and Logit

- We need a mathematical expression that connects our explanatory variables to the mean of the Bernoulli distribution
  \[ E(Y_i) = p_i, \quad 0 \leq p_i \leq 1 \]
- Two cumulative distribution functions (s-shaped) we can use to link to the Bernoulli distribution are Logit and Probit
The Link Functions

• We suppose that there is an unobserved, latent, variable $Y_i^*$ ranging from $-\infty$ to $+\infty$.

• We also suppose that $y_i^* = x_i \cdot b + u_i$, that is the latent variable is linearly related to the Xs.

• The latent variable is linked to the observed $y_i$ variable by a measurement equation
The Link Functions

\[ y_i = \begin{cases} 
1 & \text{if} & y_i^* > \tau \\
0 & \text{if} & y_i^* \leq \tau 
\end{cases} \]

• For now we assume: \( \tau = 0 \)
The Link Functions

- Recall that, \( y_i^* = x_i \beta + u_i \)

- \( \Pr(y_i = 1 \mid x_i) = \Pr(y_i^* > 0 \mid x_i) \)
  = \( \Pr(x_i \beta + u_i > 0 \mid x_i) \)
  = \( \Pr(u_i > -x_i \beta \mid x_i) \)
  = \( \Pr(u_i \leq x_i \beta \mid x_i) \)

- we need to make an assumption about the distribution of the error term
  or \( F(x_i \beta \mid x_i) \)
The Link Functions: Probit

• If the distribution is the standard normal cdf $\Phi$, we have the Probit link.

• That is,

$$\Pr(y_i = 1 \mid x_i) = \Phi(x_i b \mid x_i)$$

$$= \int_{-\infty}^{xb} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \, dt$$
The Link Functions: Logit

• If the distribution is the standard logistic distribution, we have the Logit link.

• That is,

\[ \Pr(y_i = 1 \mid x_i) = F(x_i b \mid x_i) \]

\[ = \frac{\exp(xb)}{1 + \exp(xb)} = \frac{1}{1 + \exp^{-(xb)}} \]
The Link Functions: Logit and Probit

Red Line: Probit
Blue Line: Logit
The Logit Model

• Prob of vote Rep can be expressed as:

\[
(1) \quad P_i = \frac{1}{1 + \exp^{-(xb)}}
\]

• Since \( P_i \) is non-linear in \( X \) and \( B \) we can not use OLS to estimate the parameters

\[
(6) \quad 1 - P_i = \frac{1}{1 + \exp^{(xb)}}
\]

\[
(7) \quad \frac{P_i}{1 - P_i} = \frac{1 + \exp^{(xb)}}{1 + \exp^{-(xb)}}
\]
The Logit Model

• We can make Logit Linear by taking the natural log

\[ L_i = \ln \left( \frac{P_i}{1 - P_i} \right) = xb \]

• This is called the log of the odds ratio, it is linear in X and the parameters
Features of the Logit Model

• As P goes from 0 to 1, L goes from $-\infty$ to $\infty$
• L is linear, but Probabilities are not
• Can add more regressors
• If logit is positive, odds in favor of DV=1 increase and reverse if negative
• Slope measures the change in L for a unit change in X
  – How log odds in favor of owning a house changes as income changes by 1 unit
Estimation: STATA commands

• The commands to use in STATA are similar to regress, except you type logit or probit before your DV and IV
  – Probit vote age income edu gender
  – Logit vote age income edu gender
Interpretation

• Link functions solve problem of functional form & bias, but at the expense of intuitive interpretability of the coefficients

• Recall that coefficients reflect the impact of $X$ on $Y^*$, not $Y$

• In OLS, the rate of change, or slope, only depended on values of $X$

• In Logit or Probit, we do not have a linear change
  – Depends on values of $X$ and $B$
  – And where you are on the S curve
More on Interpretation

• Logit
  – Beta coefficients tell us the change in the log-odds which accompanies a one-unit change in $X$
  – Example, if $P(Y=1)=.5$, then the odds are 1 to 1
  – If $P(Y=1)=.75$, the odds are three to 1

• Probit
  – There is no easy interpretation
Interpretation of Coefficients: What can we infer

- We can still infer the direction of the effect from the signs of the coefficients
  - As Y* increases, Pr(Y=1) increases

- We can still rely on our standard error of the coefficient to give us a t-test
  - Errors are not normally distributed, but
  - Central Limit theorem says that coefficients will be (asymptotically) normally distributed
Interpretation of Coefficients: How do we understand substantive impact?

• There is no single answer to the impact of X on Y
• We can illustrate various impacts of X on Y through a description of the predicted values
• The main idea is to hold $X_2$ to $X_n$ constant, vary $X_1$ and generate the Pr($Y=1 | X_1$)
Creating Predicted Values

• First, generate a predicted value for $Y^*$ from the model in the typical way:
  \[- E(Y^*) = xb \]

• Translate $Y^*$ into $Pr(Y=1|Y^*)$ with the link function

• If probit, $Pr(y_i = 1 \mid x_i) = \Phi(x_i b \mid x_i)$, normal distribution

\[
= \int_{-\infty}^{xb} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \, dt
\]
Predicted Values in Probit

• This means $\Pr(Y=1 \mid xb)$ is equal to the area under the normal distribution from minus infinity to $xb$
This Translates into Predictions
\[ Pr(Y=1 | xb) \]

• Which translates into predicted probabilities like this:
Creating Illustrations of the Impact of $X$

- After generating one prediction, we keep changing the value of $X_1$ (holding $X_2...X_n$ constant) and creating new predicted probabilities.
- We can plot $\Pr(Y=1)$ against $X_1$ to show the effect of $X_1$.
- Remember these effects depend on the values chosen for $X_2...X_n$.
Creating Predicted Values in Logit

• The process is exactly the same for illustrating logit models

• Only difference is the link function used to translate $Y^*$ into $\Pr(Y=1|X)$

$$\Pr(Y = 1 | X_i) = \frac{1}{1 + e^{-x_i b}}$$
More on Calculating Predicted Probabilities and Marginal Effects

• In practice, there are at least three ways you can calculate these effects
  – In Excel
  – In STATA, using CLARIFY software
  – In STATA, using spost software
  – In STATA, using “margins” command
Using CLARIFY

• Software called Clarify, runs simulations to create predicted probabilities and first differences

• You can download the software from Gary King’s web page and read the instructions

• You download it into STATA and follow the commands to use in STATA to generate the predicted probabilities and first differences
There are many ways to present substantive effects

- Clearly one could create any number of figures and tables for any given set of logit or probit results
- It is important to choose carefully
- Be clear and careful about values that other $X_2...X_n$ are set
  - Example, might want to set dummy variables at minimum or maximum
- Be clear and careful about how you vary $X_1$
Choose Meaningful Comparisons

• Pick Values that represent the data.
  – Minimum to Maximum may not be best choice
  – Don’t make predictions outside observed variation
• Choose real cases in dataset to illustrate effects
• Create hypothetical cases that correspond to meaningful values of the independent variables
Model Performance: Pseudo-$R^2$

- $R^2$ is useless for logit and probit, but there is a pseudo- $R^2$ calculated to determine the fit of the Model.
- There are MANY possible pseduo-$R^2$s computed by STATA.
  - They come up with wildly different values, really should not use them
  - [http://www.ats.ucla.edu/stat/mult_pkg/faq/general/Psuedo_RSquareds.htm](http://www.ats.ucla.edu/stat/mult_pkg/faq/general/Psuedo_RSquareds.htm)
Model Performance: Percent Correctly Predicted

• Another way to discuss the performance of the model is to report the percent of observations correctly predicted at 0 or 1

• This is easy to do in STATA

• After running the regression, type lstat
Model Performance: Proportional Reduction in Error

• How much better does our model perform compared to “dumb guessing?”

• PRE
  \[ \text{PRE} = \frac{N_{mc} - N_{nc}}{N - N_{nc}} \]
  • Where \( N_{mc} \) is the number correct under the estimated model
  • Where \( N_{nc} \) is the number correct under the null model (dummy model)
    – Best guess is that everyone will vote Republican, use actual number who did vote republican
  • Where \( N \) is total number of observations

  – Interpretation: the PRE value is that our model eliminated \( X\% \) of the error, relative to the null model
Other qualitative DV models

• Tobit Model
  – Used when you have a censored sample, that is missing $Y_i$ for some section of the population
  – Example, want to estimate amount people spend on house, but we do not have data for those who do not buy homes
  – If we estimate with only part of the sample, we end up with biased estimates and inconsistent data
Poisson Regression Model

- This model might be used if we have discrete data, or rare data
- We use this model when we think that jumping from 0 to 1 requires a steeper slope than moving from 2 to 3
More than two outcomes

• Ordinal Logit or Probit
  – More than two outcomes, on ordered dimension, ex. Ideology
  – The items are ranked, but we can not say distance between outcomes is equal

• Multinomial Logit or Probit
  – More than two outcomes, but not ordered
  – Ex. Clinton, Perot and Dole
Choosing between logit and probit