*Regression with a Binary Dependent Variable*

# *Chapter 9*

Michael Ash

CPPA



## Course Notes

- Closing remarks on quasi-experiments
- Endgame
  - Take-home final
    - Distributed in class Thursday 13 May
    - Due Tuesday 18 May (Emailed PDF ok; no Word, Excel, etc.)
  - Problem Set 7
    - Optional, worth up to 2 percentage points of extra credit
    - Due Friday 14 May
- Regression with a Binary Dependent Variable

## Econometrics of quasi-experiments

- If the same individuals are observed before and after treatment (or non-treatment), the methods are identical to experimental methods.
- More important to include $W_{1i}, \ldots, W_{ri}$ to control for pretreatment differences.
- Sometimes similar cohorts rather than the same individuals are observed over time. In this case, define...



## Econometrics of quasi-experiments

$$\begin{aligned} G_i &= 1 \text{ if } i \text{ in treatment group; 0 otherwise} \\ D_t &= 1 \text{ if } t \text{ is second period; 0 otherwise} \end{aligned}$$

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 G_i + \beta_3 D_t + \beta_4 W_{1it} + \cdots + \beta_{3+r} W_{rit} + u_i t$$

where

$$X_{it} = G_i \times D_t$$

and $\hat{\beta}_1$ is the d-i-d estimate of the causal effect.

# Problems with Quasi-Experiments

- Failure of randomization (more likely than in an experiment). Can perform the same test *on observables*.
- Failure to follow treatment protocol
- Attrition
- Hawthorne effects are not germane for a quasi-experiment
- “Instrument validity in quasi-experiments”: does the effect of the quasi-experimental assignment leak through (affect $Y$) other than in the hypothesized way.
- External Validity: risks may be greater because the institutional circumstances are quirky and unchosen.



# Heterogeneous Populations

(Quasi-)Experimental Estimates in Heterogeneous Populations

$$Y_i = \beta_0 + \beta_{1\mathbf{i}} X_i + u_i$$

Whoops! The effect of $X$ differs for each person.

- $\hat{\beta}_1^{OLS}$ captures the average causal effect, $E(\beta_{1\mathbf{i}})$
- $E(\beta_{1\mathbf{i}})$ is a useful policy variable (although worthless for case workers)
- If the sub-populations for which $\beta_{i1}$ differs can be identified, results can be stratified (as in MTO).
- Quasi-experiments may falter because they identify the average $\beta_{i1}$ for the individuals whom the quasi-experiment assigned.

## Closing thoughts on experiments

- Experiment is a useful benchmark for observational studies
- Use the experiment to focus on how an observational study might fail.
- Quasi-experiments are feasible, even on small scales.



## Home Mortgage Lending Discrimination

- Loan candidates identical but for race
- Discrimination in loan application approval.
- Discrimination in value, interest rates, points, discouragement.
- Redlining (race of neighborhood, not necessarily race of applicant.)
- Direct and indirect (education, employment, health) pernicious consequences
- Boston: 28 percent denial rate for black applicants, 9 percent for white.
  - Other factors? Omitted variables? Need multivariate analysis.
- HMDA Data: a powerful tool for credit justice advocacy

# Binary Dependent Variables

- Outcome can be coded 1 or 0 (yes or no, approved or denied, success or failure) Examples?
- First approaches cross-tabulation and $t$-test for proportions
  - But what if the explanatory variable is continuous, e.g., $P/I$ ratio (Figure 9.1)?
- Interpret the regression as modeling the probability that the dependent variable equals one ($Y = 1$).
  - For a binary variable, $E(Y) = \Pr(Y = 1)$
- Multiple explanatory variables? Stratify or multivariate methods



# Probability and Changes in Probability

**Probability:** $\Pr(Y = 1)$ proportion of the time $[0, 1]$ that an outcome occurs, in the long run.

**Change in probability for a change in $X$:** $\frac{\Delta \Pr(Y=1)}{\Delta X}$ change in probability

**Odds:** $\frac{p}{1-p}$, the ratio of the probability that an event occurs to the the probability that it does not occur (note that odds are close to probability if the probability is low)

**Odds Ratio for a change in $X$:** $\frac{(\frac{p}{1-p})|X+\Delta X}{(\frac{p}{1-p})|X}$

# HMDA example

- Outcome: loan denial is coded $1$, loan approval $0$
- Key explanatory variable: black
- Other explanatory variables: $P/I$, credit history, LTV, etc.



# Linear Probability Model (LPM)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$$

Simply run the OLS regression with binary $Y$.

- $\beta_1$ expresses the change in probability that $Y = 1$ associated with a unit change in $X_1$.
- $\hat{Y}_i$ expresses the probability that $Y_i = 1$

$$\Pr(Y = 1 | X_1, X_2, \ldots, X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k = \hat{Y}$$

# HMDA Example

$$\widehat{\text{deny}} = \underset{(0.032)}{-0.080} + \underset{(0.098)}{0.604}\, P/I$$

See Figure 9.1 (a scatterplot)

$$\widehat{\text{deny}} = \underset{(0.029)}{-0.091} + \underset{(0.089)}{0.559}\, P/I + \underset{(0.025)}{0.177}\, \text{black}$$

Interpretation of $\beta_{\text{black}}$ and $\beta_{P/I}$. Significance.



# Shortcomings of the LPM

- “Nonconforming Predicted Probabilities” Probabilities must logically be between 0 and 1, but the LPM can predict probabilities outside this range.
- Heteroskedastic by construction (always use robust standard errors)
- Excellent starting place

## Probit and Logit Regression

- Addresses nonconforming predicted probabilities in the LPM
- Basic strategy: bound predicted values between $0$ and $1$ by transforming a linear index, $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$, which can ranges over $(-\infty, \infty)$ into something that ranges over $[0, 1]$
- When the index is big and positive, $\Pr(Y = 1) \to 1$.
- When the index is big and negative, $\Pr(Y = 1) \to 0$.
- How to transform? Use a Cumulative Distribution Function.



## Probit Regression

The CDF is the cumulative standard normal distribution, $\Phi$.
The index $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$ is treated as a $z$-score.

$$\Pr(Y = 1 | X_1, X_2, \ldots, X_k) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)$$

## Interpreting the results

$$\Pr(Y=1|X_1,X_2,\ldots,X_k)=\Phi(\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_k X_k)$$

- $\beta_j$ positive (negative) means that an increase in $X_j$ increases (decreases) the probability of $Y=1$.
- $\beta_j$ reports how the *index* changes with a change in $X$, but the index is only an input to the CDF.
- The size of $\beta_j$ is hard to interpret because the change in probability for a change in $X_j$ is non-linear, depends on all $X_1,X_2,\ldots,X_k$.
- Easiest approach to interpretation is computing the predicted probability $\hat{Y}$ for alternative values of $X$
- Same interpretation of standard errors, hypothesis tests, and confidence intervals as with OLS



## HMDA example

See Figure 9.2

$$\widehat{\Pr(\text{deny}=1|P/I,\text{black})}=\Phi(\underset{(0.16)}{-2.26}+\underset{(0.44)}{2.74}\,P/I+\underset{(0.083)}{0.71}\,\text{black})$$

- White applicant with $P/I=0.3$: $\widehat{\Pr(\text{deny}=1|P/I,\text{black})}=\Phi(-2.26+2.74\times 0.3+0.71\times 0)=\Phi(-1.44)=7.5\%$
- Black applicant with $P/I=0.3$: $\widehat{\Pr(\text{deny}=1|P/I,\text{black})}=\Phi(-2.26+2.74\times 0.3+0.71\times 1)=\Phi(-0.71)=23.3\%$

## Logit Regression

Logit uses a slightly different functional form of the CDF (the logistic function) instead of the standard normal CDF.
The coefficients of the index can look different, but the probability results are usually very similar to the results from probit and from the LPM.
Aside from the problem with non-conforming probabilities in the LPM, the three models generate similar results.



## Estimation and Inference in Logit and Probit Models

- OLS (and LPM, which is an application of OLS) has a closed form formula for $\hat{\beta}$
- Logit and Probit require numerical methods to find $\hat{\beta}$ that best fits the data.

# Other LDV Models

Limited Dependent Variable (LDV)

- Censored and Truncated Regression Models. Tobit or sample selection models.
- Count Data (discrete non-negative integers), $Y \in 0, 1, 2, \ldots, k$ with $k$ small. Poisson or negative binomial regression.
- Ordered Responses, e.g., completed educational credentials. Ordered logit or probit.
- Discrete Choice Data, e.g., mode of travel. Characteristics of choice, chooser, and interaction. Multinomial logit or probit,
- Can sometimes convert to several binary problems.