Regression with a Binary Dependent Variable

Chapter 9

Michael Ash

CPPA

Lecture 22

Course Notes

- Endgame
  - Take-home final
    - Distributed Friday 19 May
    - Due Tuesday 23 May (Paper or emailed PDF ok; no Word, Excel, etc.)
  - Problem Set 7
    - Optional, worth up to 2 percentage points of extra credit
    - Due Friday 19 May
- Regression with a Binary Dependent Variable

Binary Dependent Variables

- Outcome can be coded 1 or 0 (yes or no, approved or denied, success or failure) Examples?
- Interpret the regression as modeling the probability that the dependent variable equals one ($Y = 1$).
  - Recall that for a binary variable, $E(Y) = \Pr(Y = 1)$

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.

- β_1 expresses the change in probability that $Y = 1$ associated with a unit change in X_1.
- $\hat{Y}_i$ expresses the probability that $Y_i = 1$

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

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)

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 range over $(-\infty, \infty)$ into something that ranges over $[0, 1]$
- When the index is big and positive, $\Pr(Y = 1) \rightarrow 1$.
- When the index is big and negative, $\Pr(Y = 1) \rightarrow 0$.
- How to transform? Use a Cumulative Distribution Function.

Probit Regression

The CDF is the cumulative standard normal distribution, Φ.
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

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

- β_j positive (negative) means that an increase in X_j increases (decreases) the probability of $Y = 1$.
- β_j reports how the *index* changes with a change in X, but the index is only an input to the CDF.
- The size of β_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 or Logistic Regression

Logit, or logistic regression, 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 predicted probabilities.

Estimation of 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 $\hat{\beta}$'s that best fit the data.

Nonlinear Least Squares

One approach is to choose coefficients $b_0, b_1, \ldots, b_k$ that minimize the sum of squares of how far the actual outcome, Y_i, is from the prediction, $\Phi(b_0 + b_1 X_{1i} + \cdots + b_k X_{ki})$.

$$\sum_{i=i}^{n} [Y_i - \Phi(b_0 + b_1 X_{1i} + \cdots + b_k X_{ki})]^2$$

Maximum Likelihood Estimation

- An alternative approach is to choose coefficients $b_0, b_1, \ldots, b_k$ that make the current sample, $Y_1, \ldots, Y_n$ as likely as possible to have occurred.
- For example, if you observe data $\{4, 6, 8\}$, the predicted mean that would make this sample most likely to occur is $\hat{\mu}^{MLE} = 6$.
- Stata probit and logistic regression (logit) commands are under Statistics $\rightarrow$ Binary Outcomes

Inference and Measures of Fit

- Standard errors, hypothesis tests, and confidence intervals are exactly as in OLS, **but** they refer to the coefficients and must be translated into probabilities by applying the appropriate CDF.

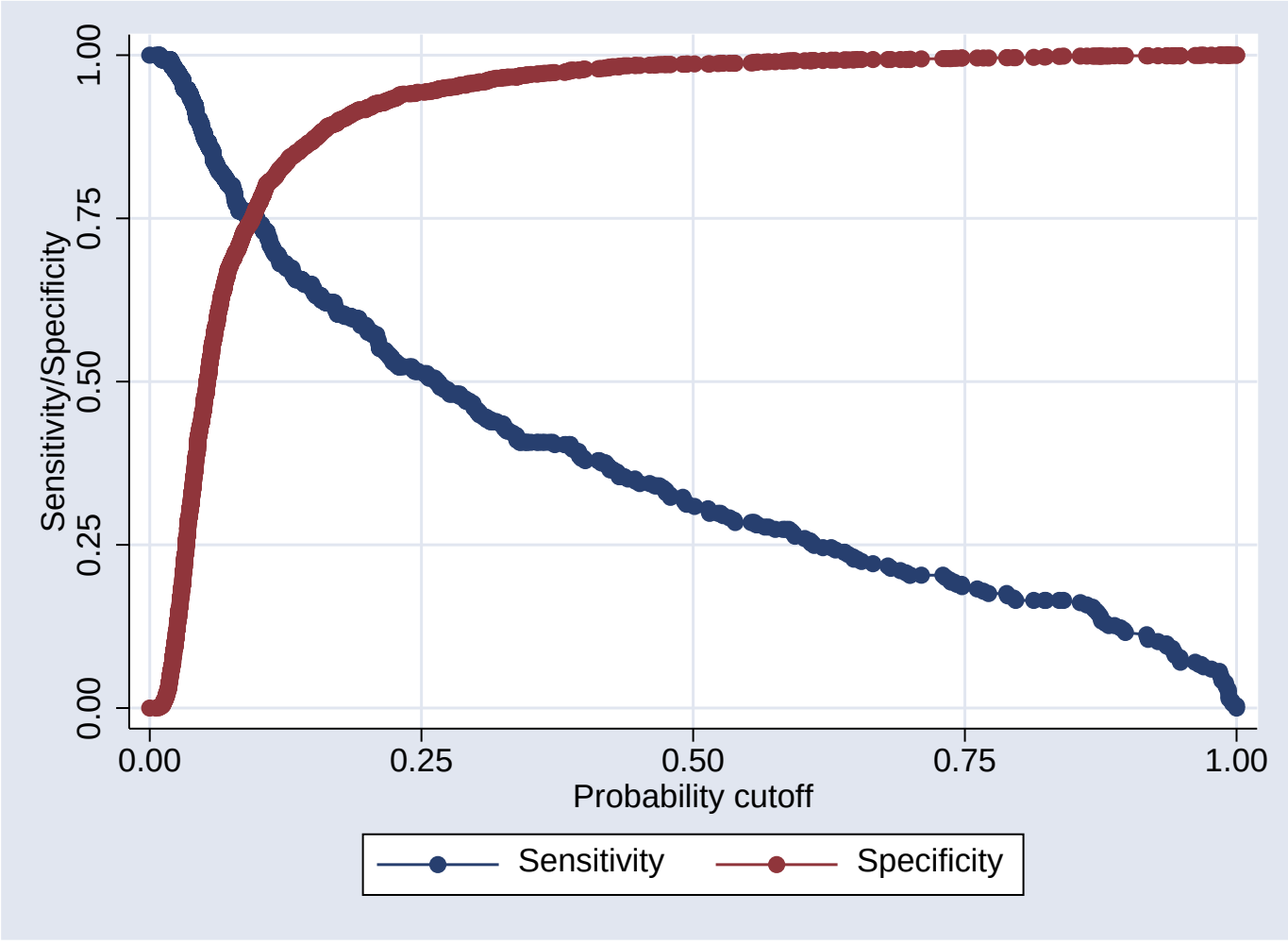
fraction correctly predicted using one probability cut-off, e.g., 0.50, and check the fraction correctly predicted, but. . .

sensitivity/specificity Choose a cut-off. Sensitivity is the fraction of observed positive-outcomes that are correctly classified. Specificity is the fraction of observed negative outcomes that are correctly specified.

Pseudo-R^2 is analogous to the R^2

- Expresses the predictive quality of the model with explanatory variables relative to the predictive quality of the sample proportion p of cases where $Y_i = 1$
- Adjusts for adding extra regressors

Sensitivity and Specificity



Reviewing the HMDA results (Table 9.2)

- LPM, logit, probit (minor differences)
- Four probit specifications
- Highly robust result: 6.0 to 8.4 percentage-point gap in white-black denial rates, controlling for a wide range of other explanatory variables.
- Internal Validity
- External Validity

Other LDV Models

Limited Dependent Variable (LDV)

- 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.
- Censored and Truncated Regression Models. Tobit or sample selection models.