Binary-Valued Dependent Variables

Ass.Prof. Andriy Stavytskyy

Agenda

- Binary-Valued Dependent Variables
- Probit/Logit Models
- Estimating a Probit/Logit Model
- Deriving Probit/Logit

BINARY-VALUED DEPENDENT VARIABLES

Binary Dependent Variables - 1

- We have worked extensively with regression models in which Y is continuous.
  - *We have predicted the effect of education and experience on earnings.*
  - *We have predicted the effect of exogenous changes in price on quantity demanded.*

Binary Dependent Variables – 2

- However, our methods are inappropriate when the dependent variable takes on just a few discrete values.
  - *For example, we may be interested in the effect of a brand's advertising on consumers' decisions to buy that brand.*

Why Do We Need A Different Model Than Linear Regression?

Appropriate estimation of relations between variables depends on selecting an appropriate statistical model. There are many different types of estimation problems in political science.

- ***Continuous variables*** where the experiment can be viewed as draws from a normal distribution.
- ***Continuous Variables*** where the distribution is truncated or censored.
- ***Discrete Variables*** - for example, we might model labor force participation, whether to vote for or against, purchase or not purchase, run for office or not run for office, etc.

Type of Qualitative Response Models

- ***Qualitative dichotomy*** (e.g., vote/not vote type variables)- We equate "no" with zero and "yes" with 1. However, these are qualitative choices and the coding of 0-1 is arbitrary. We could equally well code "no" as 1 and "yes" as zero.
- ***Qualitative multichotomy*** (e.g., occupational choice by an individual)- Let 0 be a clerk, 1 an engineer, 2 an attorney, 3 a politician, 4 a college professor, and 5 other. Here the codings are mere categories and the numbers have no real meaning.
- ***Rankings*** (e.g., opinions about a politician's job performance)- Strongly approve (5), approve (4), don't know (3), disapprove (2), strongly disapprove (1). The values that are chosen are not quantitative, but merely an ordering of preferences or opinions. The difference between outcomes is not necessarily the same from 5 to 4 as it is from 2 to 1.
- ***Count outcomes***.

Binary Dependent Variables - 3

- ***<u>Discrete-valued</u>*** dependent variables are a special case that comes up sufficiently frequently to warrant its own special techniques.
- Here we will focus on dependent variables that can take on only 2 values, 0 or 1 (dummy variables).

Example

Suppose we were to predict whether football teams win individual games, using the reported point spread from sports gambling authorities.

Example: model – 1

- Using the techniques we have developed so far, we might regress

$$D_i^{Win} = \beta_0 + \beta_1 Spread_i + \varepsilon_i$$

where i indexes games

- *How would we interpret the coefficients and predicted values from such a model?*

Example: model - 2

$$D_i^{Win} = \beta_0 + \beta_1 Spread_i + \varepsilon_i$$

- D_i^{Win} is either 0 or 1. It does not make sense to say that a 1 point increase in the spread increases D_i^{Win} by b1. D_i^{Win} can change only from 0 to 1 or from 1 to 0.
- Instead of predicting D_i^{Win} itself, we predict the probability that $D_i^{Win} = 1$.

Binary Dependent Variables – 4

$$D_i^{Win} = \beta_0 + \beta_1 Spread_i + \varepsilon_i$$

- It can make sense to say that a 1 point increase in the spread increases the probability of winning by β_1.
- Our predicted values of D_i^{Win} are the probability of winning.

Binary Dependent Variables – 5

$$D_i^{Win} = \beta_0 + \beta_1 Spread_i + \varepsilon_i$$

When we use a linear regression model to estimate probabilities, we call the model the linear probability model.

Problems with LPM Regression

- OLS in this case is called the Linear Probability Model
- Running regression produces some problems
  - *Errors are not distributed normally*
  - *Errors are heteroskedastic*
  - *Predicted Ys can be outside the 0.0-1.0 bounds required for probability*

What Point Spreads Say About the Probability of Winning in Football?

Dependent Variable: WIN
Method: Least Squares
Sample: 1 644
Included observations: 644
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.500000	0.018849	26.52593	0.0000
SPREAD	−0.025180	0.003068	−8.26065	0.0000

R-squared	0.087582	Mean dependent var	0.500000
Adjusted R-squared	0.086161	S.D. dependent var	0.500389
S.E. of regression	0.478346	Akaike info criterion	1.366137
Sum squared resid	146.8993	Schwarz criterion	1.380012
Log likelihood	−437.8960	*F*-statistic	61.62496
Durbin–Watson stat	2.034242	Prob(*F*-statistic)	0.000000

Model charactersistics

- Note that the table reports ***White Robust Estimated Standard Errors***.
- The Linear Probability Model disturbances are ***heteroskedastic***.
- Heteroskedasticity is the only violation of the Gauss–Markov assumptions inherent in using dummy variables as Y.

Model analysis - 1

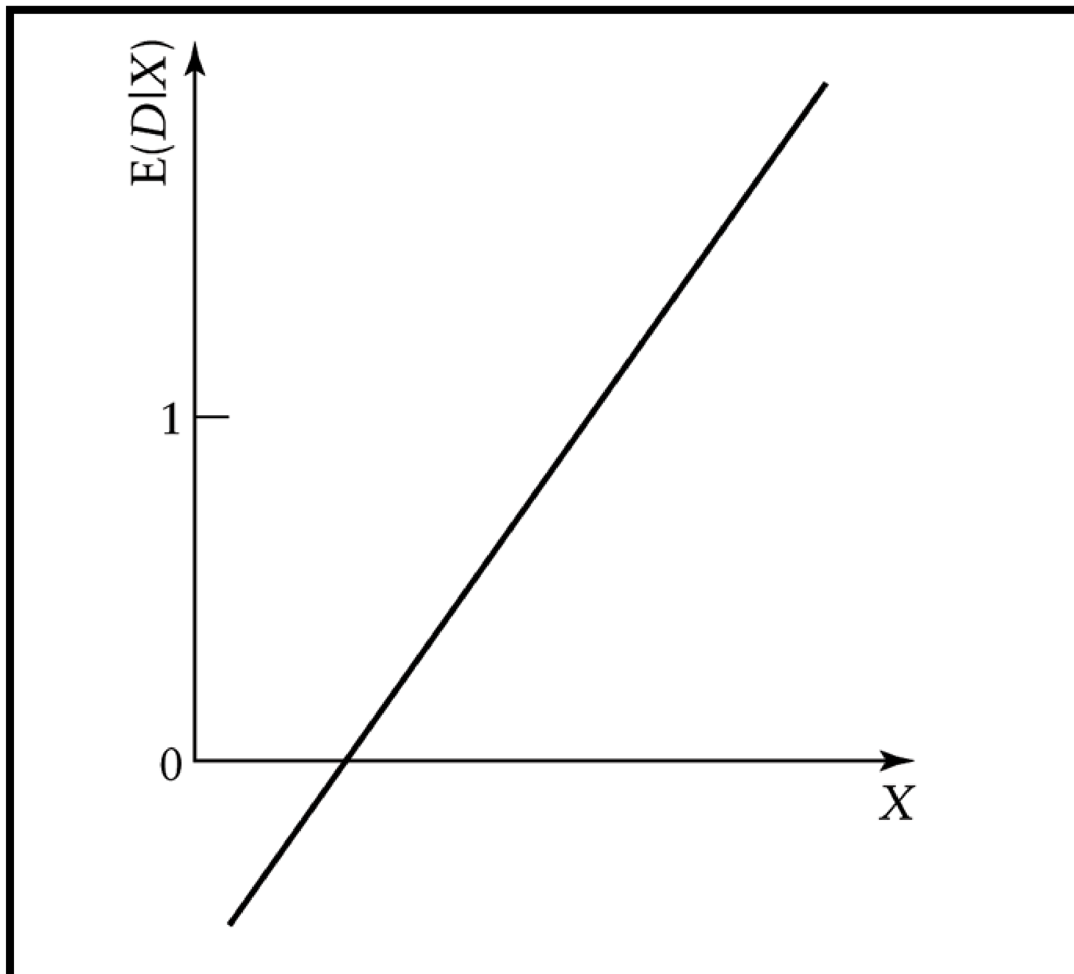
- The linear probability model works fine mathematically.
- However, it faces a serious drawback in interpretation.
- If the point spread is 21 points, the team's predicted probability of winning is:

$$0.5 - 0.025 \cdot 21 = -0.025$$

Model analysis – 2

- If X = 21, E(Y | X) = -0.025
- We predict that the team has a -2.5% probability of victory.
- If X = -21, we predict that the team has a 102.5% probability of victory.

For Some X-Values, E(D|Xi) > 1
For Some Other Values E(D|Xi) < 0



Requirements

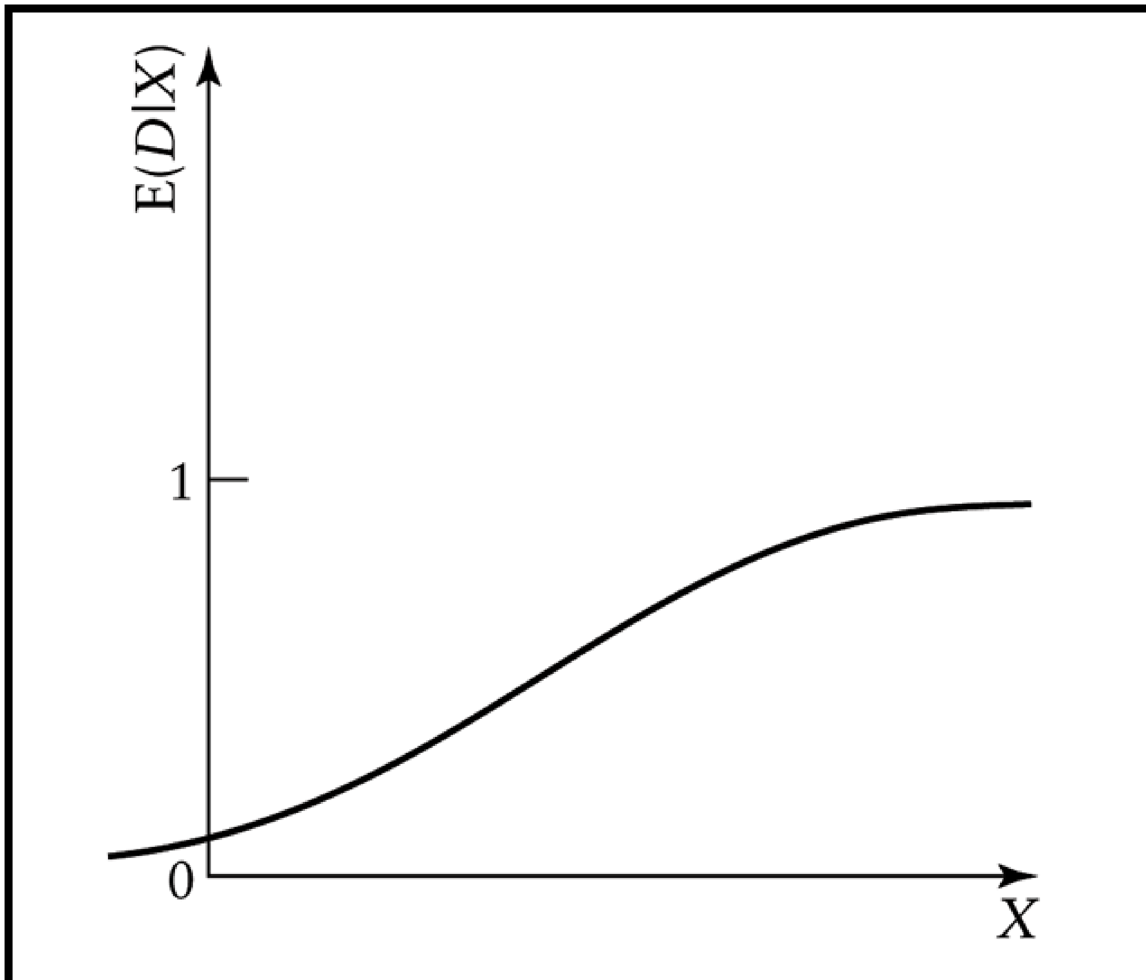
- Linear regression methods predict values between $-\infty$ and $+\infty$.
- Probabilities must fall between 0 and 1.
- The linear probability model cannot guarantee sensible predictions.

Translator

<u>We want a translator such that</u>:

- The closer to -∞ is the value from our linear regression model, the closer to 0 is our predicted probability.
- The closer to +∞ is the value from our linear regression model, the closer to 1 is our predicted probability.
- No predicted probabilities are less than 0 or greater than 1.

A Graph of Probability of Success and X



Questions

- How can we construct such a translator?
- How can we estimate it?

PROBIT/LOGIT MODELS

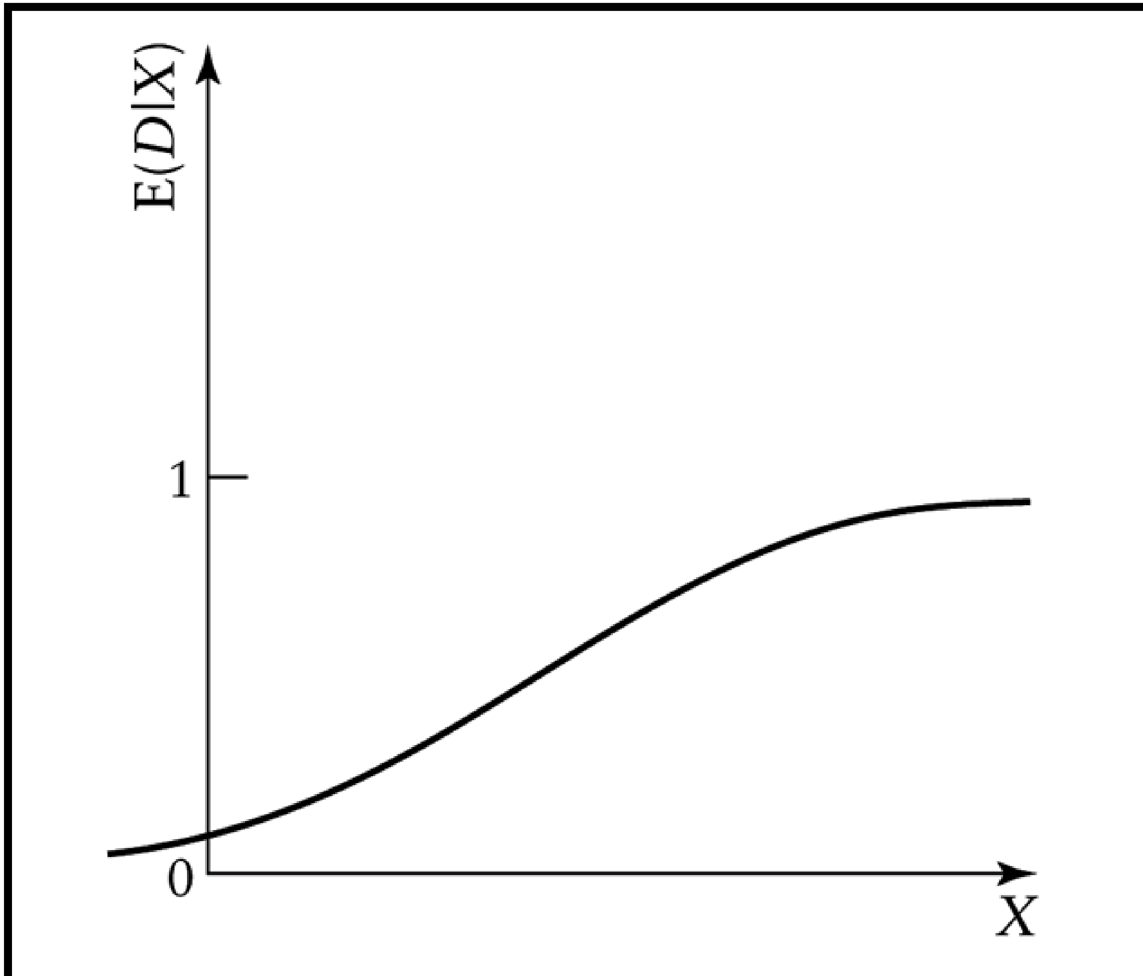
Probit/Logit Models

- In common practice, econometricians use THREE such “translators”:
  - *probit*
  - *logit*
  - *gompit*
- The differences between these models are subtle.
- For present purposes there is no practical difference between the models.

Prepositions

- Notice that the slope varies dramatically.
- When the team is very - very likely or very - very unlikely to win, a small change in the point spread has very little impact.
- When the team's chance of victory is 50/50, a small change in the point spread can lead to a large change in probabilities.

A Graph of Probability of Success and X



Structure of Probit/Logit Models - 1

Both the ***Probit*** and ***Logit*** models have the same basic structure.

- *Estimate a latent variable Z using a linear model. Z ranges from negative infinity to positive infinity.*
- *Use a non-linear function to transform Z into a predicted Y value between 0 and 1.*

Structure of Probit/Logit Model – 2

- Suppose there is some unobserved continuous ***variable Z*** that can take on values from negative infinity to infinity.
- ***The higher E(Z)*** is, the more probable it is that a team will win, or a student will graduate, or a consumer will purchase a particular brand.

Latent variable – 1

We call an unobserved variable, Z, that we use for intermediate calculations, a ***latent variable.***

Latent variable – 2

- **Z** is a linear function of the explanators:
- Our goal is to estimate these β_{i}'s.

$$Z = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki} + \varepsilon_i$$

Latent variable – 3

- We will focus particularly on E(Z):

$$E(Z) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki}$$

- It is convenient to consider the E(Z) separately from its stochastic component.

Probit/Logit/Gompit difference

The predicted probability of Y is a non-linear function of E(Z).

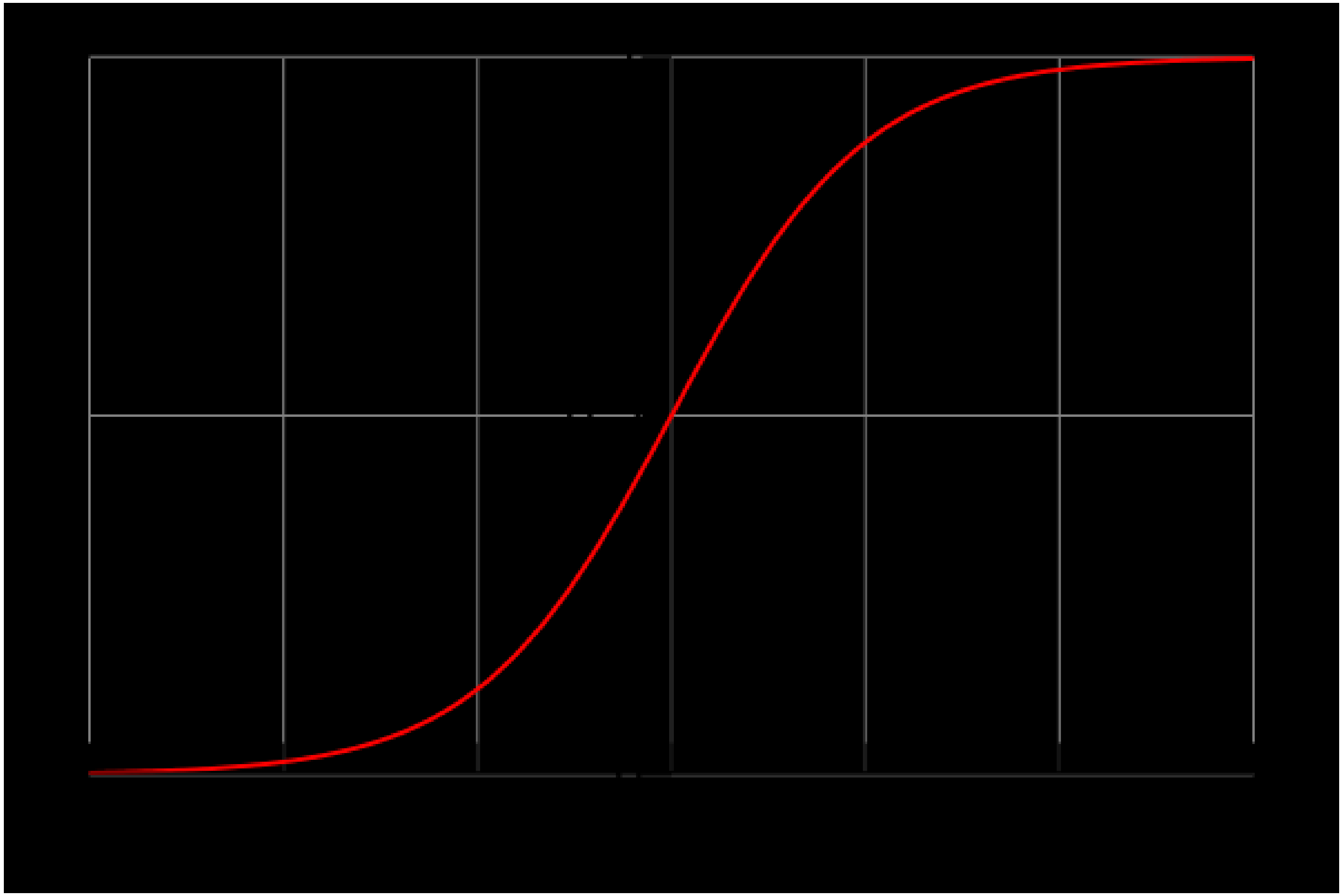
- *The **probit** model uses the standard normal cumulative density function.*
- *The **logit** model uses the logistic cumulative density function.*
- *The **gompit** model is based upon the CDF for the Type-I extreme value distribution. Note that this distribution is skewed.*

Logistic Model

- We need a model that produces true probabilities
- The Logit, or cumulative logistic distribution offers one approach.
- This produces a sigmoid curve.
- Look at equation under 2 conditions:
  - $X_i = +\infty$
  - $X_i = -\infty$

$$Y_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_i)}}$$

Sigmoid curve



Probability Ratio

- Note that

$$P_i = \frac{1}{1+e^{-(\beta_0+\beta_1 X_i)}} = \frac{1}{1+e^{-Z_i}} = \frac{e^Z}{1+e^Z}$$

- Where

$$Z_i = \beta_0 + \beta_1 X_i$$

Log Odds Ratio

- The logit is the log of the odds ratio, and is given by:

$$L_i = \ln\left(\frac{P_i}{1-P_i}\right) = Z_i = \beta_0 + \beta_1 X_i$$

- This model gives us a coefficient that may be interpreted as a change in the weighted odds of the dependent variable

Estimation of Model

- We estimate this with ***maximum likelihood***
- The significance tests are ***z statistics***
- We can generate a ***Pseudo R^2*** which is an attempt to measure the percent of variation of the underlying logit function explained by the independent variables
- We test the full model with the ***Likelihood Ratio test (LR)***, which has a χ^2 distribution with k degrees of freedom

Probit

- If we can assume that the dependent variable is actually the result of an underlying (and immeasurable) propensity or utility, we can use the cumulative normal probability function to estimate a Probit model
- Also, more appropriate if the categories (or their propensities) are likely to be normally distributed
- It looks just like a logit model in practice

The Cumulative Normal Density Function

- The normal distribution is given by:

$$f(X) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\left(\frac{(X-\mu)^2}{2\sigma^2}\right)}$$

- The Cumulative Normal Density Function is:

$$F(X) = \int_{-\infty}^{X_0} \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\left(\frac{(X-\mu)^2}{2\sigma^2}\right)}$$

The Standard Normal CDF

- We assume that there is an underlying threshold value (I_i) that if the case exceeds will be a 1, and 0 otherwise.
- We can standardize and estimate this as

$$F(I_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_0 + \beta_1 X_i} e^{-z^2/2} dz$$

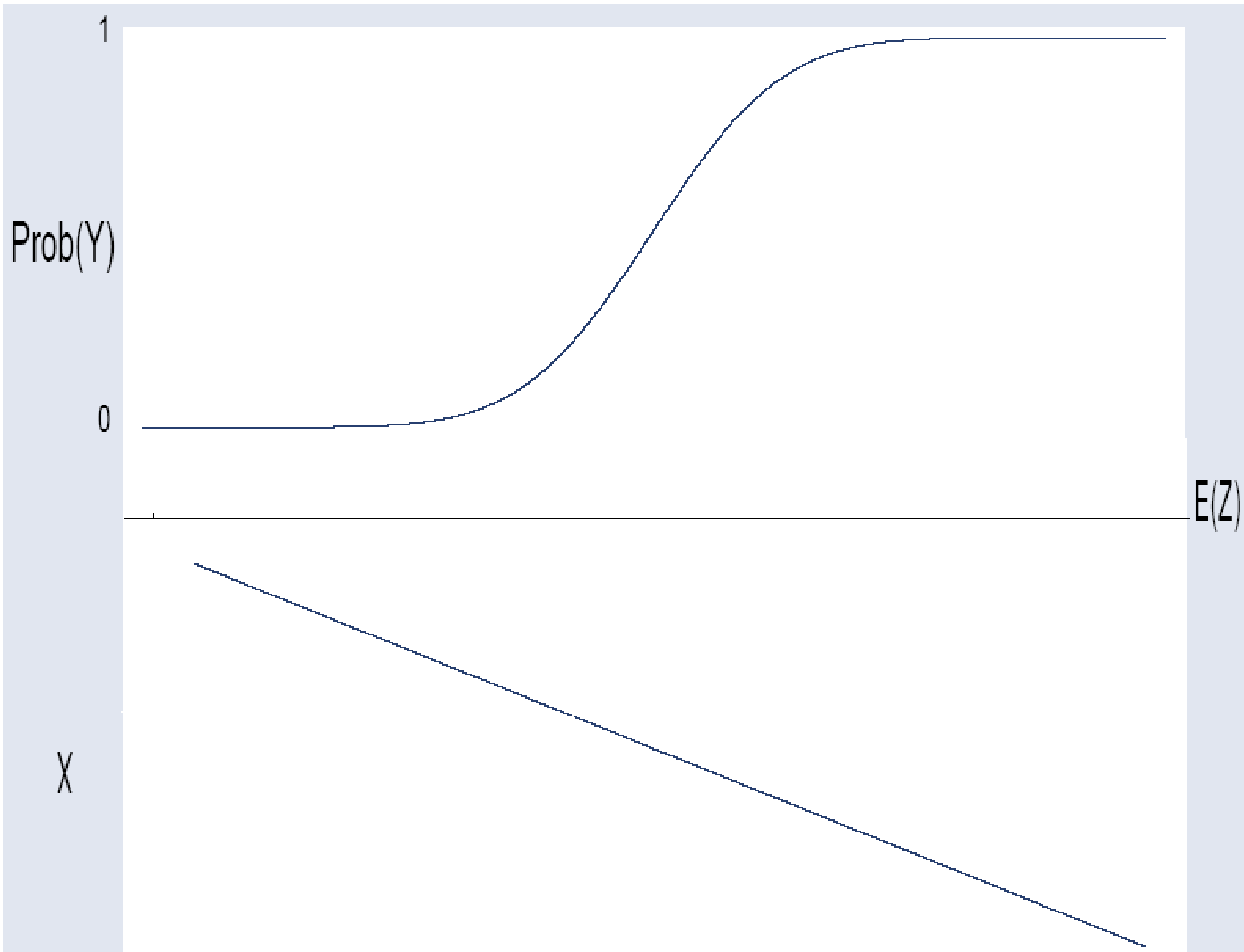
Probit estimates

- Again, maximum likelihood estimation
- Again, a Pseudo R^2
- Again, a LR ratio with k degrees of freedom

Assumptions of Models

- All Y's are in {0,1} set
- They are statistically independent
- No multicollinearity
- The $P(Y_i=1)$ is normal density for probit, and logistic function for logit

Graph

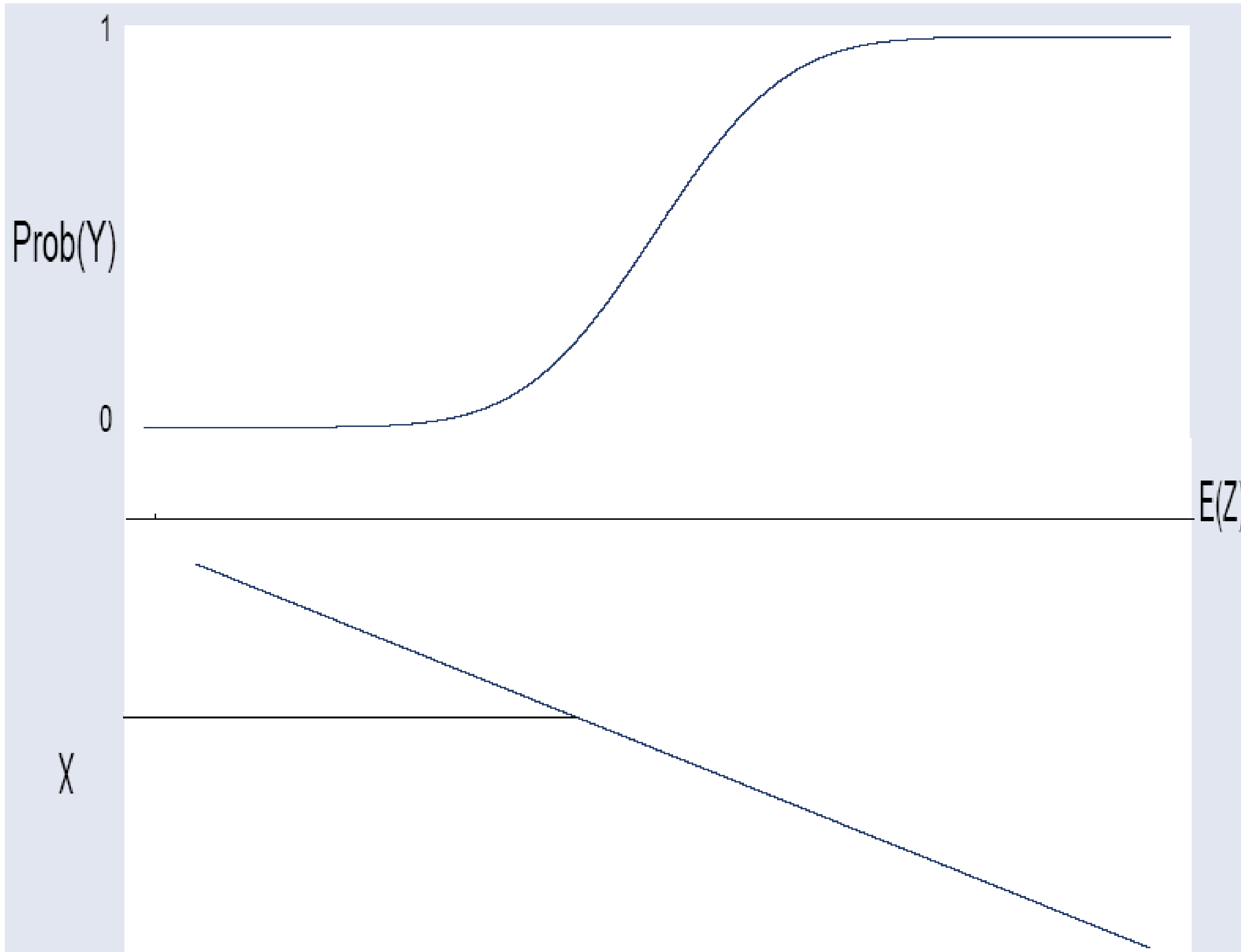


Prediction

- To predict the Prob(Y) for a given X value, begin by calculating the fitted Z value from the predicted linear coefficients.
- For example, if there is only one explanator X:

$$E(Z) = \hat{Z}_i = \beta_0 + \beta_1 \hat{X}_i$$

Graph-prediction – 1

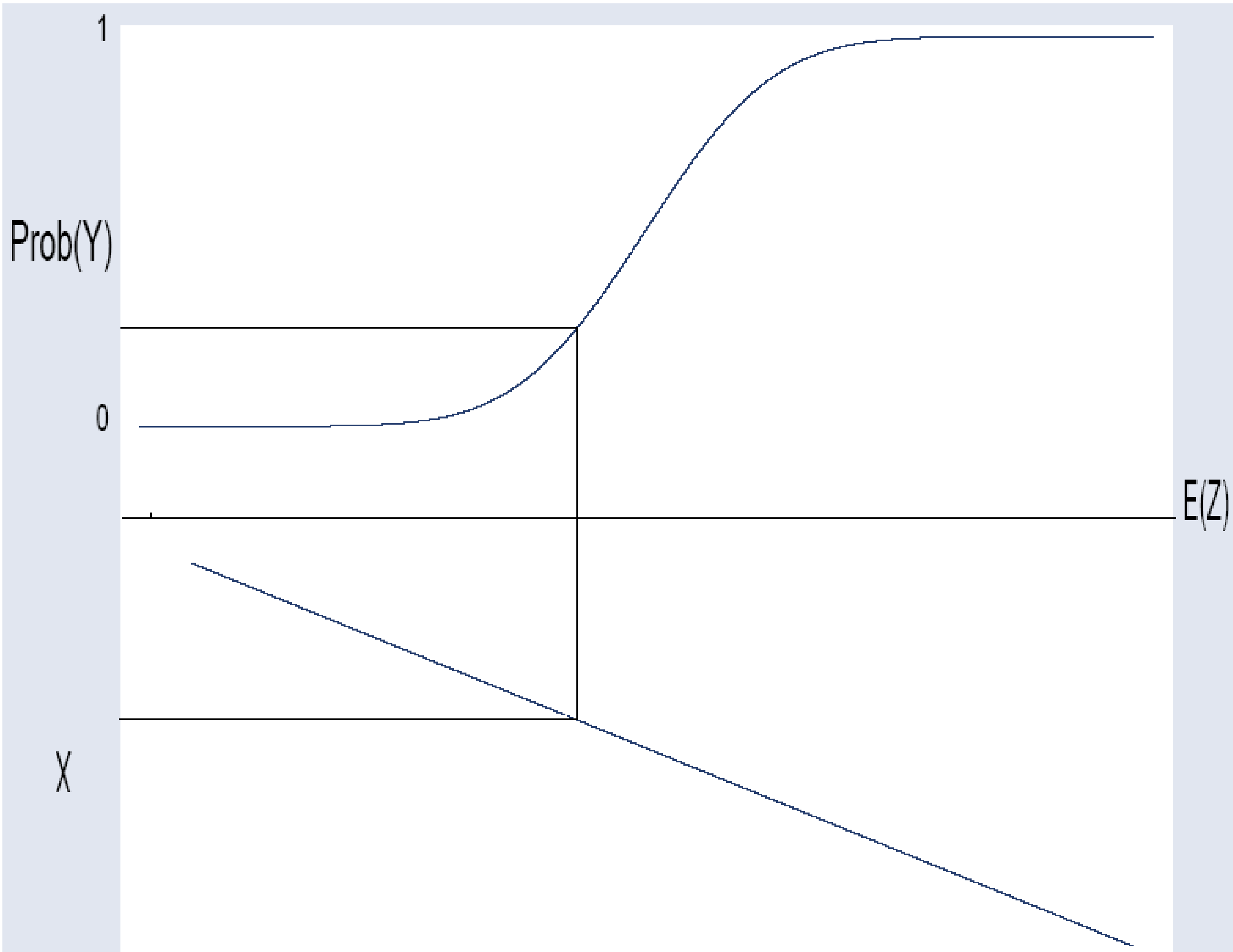


Probit/Logit Model prediction

- Then use the nonlinear function to translate the fitted Z value into a Prob(Y):

$$Prob(Y) = F(\hat{Z})$$

Graph-prediction – 2



ESTIMATING A PROBIT/LOGIT MODEL

Estimating a Probit/Logit Model

Each model is estimated using a statistical method called the method of **maximum likelihood.**

Estimating a Probit/Logit Model – 2

You must specify three elements:

- The dummy outcome variable (whether the football team actually won game i)
- The explanator/s (the football team's point spread for game i)
- Which nonlinear function F(•) you wish to use (you specify F when you tell the computer whether to use logit or probit)

Estimating a Probit/Logit Model – 2

- The computer then calculates the β_i's that assigns the highest probability to the outcomes that were observed.
- The computer can calculate the β_i's for you. You must know how to interpret them.

Model estimation

Dependent Variable: WIN
Method: ML - Binary Logit (Quadratic hill climbing)
Time: 17:21
Sample: 1 644
Included observations: 644
Convergence achieved after 3 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	6.55E-17	0.082525	7.94E-16	1.0000
SPREAD	−0.109814	0.015211	−7.219175	0.0000

Mean dependent var	0.500000	S.D. dependent var	0.500389
S.E. of regression	0.477754	Akaike info criterion	1.301076
Sum squared resid	146.5357	Schwarz criterion	1.314951
Log likelihood	−416.9466	Hannan-Quinn criter.	1.306460
Restr. log likelihood	−446.3868	Avg. log likelihood	−0.647433
LR statistic (1 df)	58.88031	McFadden R-squared	0.065952
Probability(LR stat)	1.68E-14		
Obs with Dep = 0	322	Total obs	644
Obs with Dep = 1	322		

Analysis of Probit/Logit Model

- The estimated slope of the point spread is -0.1098
- A 1-point increase in the point spread decreases E(Z) by 0.1098 units.
- How do we interpret the slope dZ/dX ?

Analysis: Statistical significance

You can still read statistical significance from the slope **dZ/dX**. The z-statistic reported for probit or logit is analogous to OLS's t-statistic.

Analysis: Sign

- ***<u>Sign:</u>*** if dZ/dX is positive, then dProb(Y)/dX is also positive.
- The z-statistic on the point spread is -7.22, well exceeding the 5% critical value of 1.96. The point spread is a statistically significant explanator of winning football games.
- The sign of the coefficient is negative. A higher point spread predicts a lower chance of winning.

Analysis: Magnitude

- the magnitude of ***dZ/dX*** has no particular interpretation. We care about the magnitude of ***dProb(Y)/dX***.
- From the computer output for a probit/logit estimation, you can interpret the statistical significance and sign of each coefficient directly. Assessing magnitude is trickier.

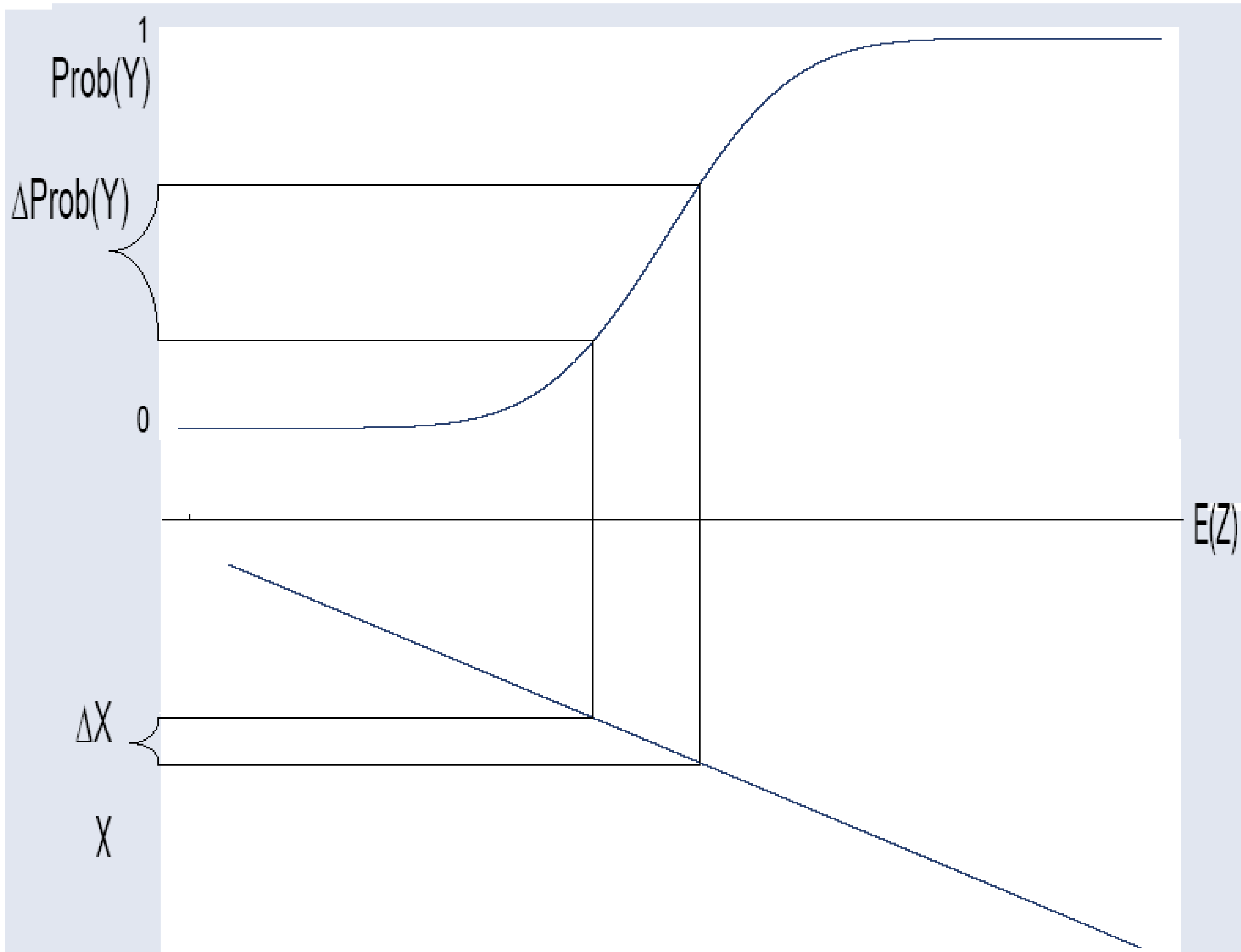
Problems in Interpreting Magnitude

- The estimated coefficient relates X to Z. We care about the relationship between X and Prob(Y = 1).
- The effect of X on Prob(Y = 1) varies depending on Z.

First approach to assessing the magnitude

1. One approach is to predict Prob(Y) for different values of X, to see how the probability changes as X changes.

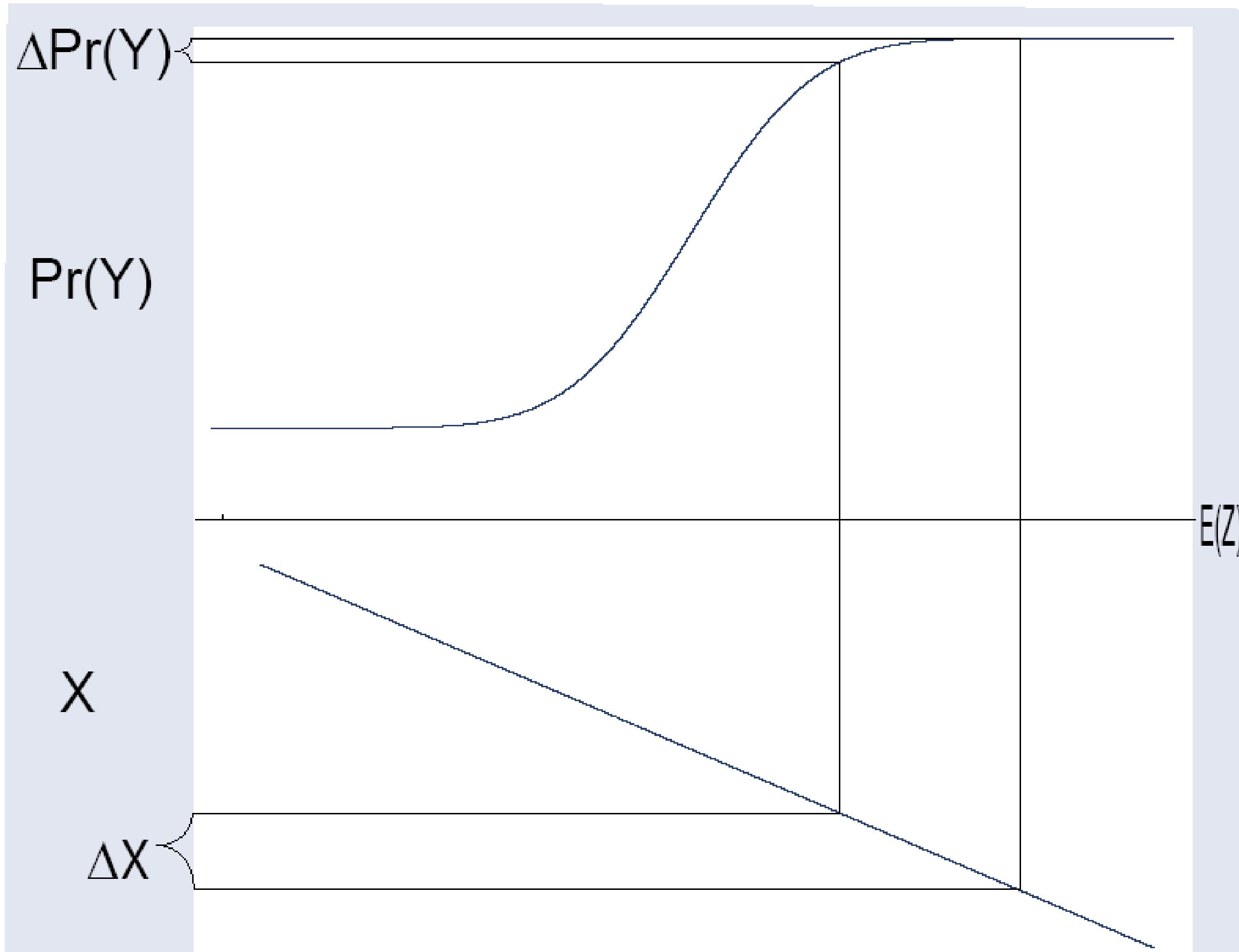
Graph – 1



But…

- the effect of a 1-unit change in X varies greatly, depending on the initial value of E(Z).
- E(Z) depends on the values of all explanators.

Graph – 2



Example - 1

- For example, let's consider the effect of 1 point change in the point spread at SPREAD = 5.88 points.
- Note: In this example, there is only one explanator, SPREAD. If we had other explanators, we would have to specify their values for this calculation, as well.

Example – 2

- Step One: Calculate the E(Z) values for X = 5.88 and X = 6.88, using the fitted values.
- Step Two: Plug the E(Z) values into the formula for the logistic density function.

Example – 3

$$Z(5.88) = 0 - 0.1098 \bullet 5.88 = 0.6456$$

$$Z(6.88) = 0 - 0.1098 \bullet 6.88 = 0.7554$$

For the logit, $F(\hat{Z}) = \dfrac{\exp(\hat{Z})}{1 + \exp(\hat{Z})}$

$$F(0.7554) - F(0.6456) = 0.6560 - 0.6804 = -0.0243.$$

Example - 4

- Changing the point spread from 5.88 to 6.88 predicts a 2.4 percentage point decrease in the team's chance of victory.
- Note that changing the point spread from 8.88 to 9.88 predicts only a 2.1 percentage point decrease.

Second approach to assessing the magnitude

$$\frac{dProb(Y)}{dX_1} = \frac{dProb(Y)}{d\hat{Z}} \bullet \frac{d\hat{Z}}{X_1} = \frac{dF}{d\hat{Z}} \bullet \hat{\beta}_1$$

Unfortunately, $\frac{dF}{d\hat{Z}}$ varies, depending on $\hat{Z}$. However, a sample value can be calculated for a representative $\hat{Z}$ value. Typically, we use the $\hat{Z}$ calculated at the mean values for each X.

But...

- Some econometrics software packages can calculate such "***pseudo-slopes***" for you.
- EViews does NOT have this function.

Example

The following table reports a probit on the probability of holding interest-bearing assets, as a function of total financial assets (LNFINAST) and dummy variables for having a pension (PENSION) or IRA (IRAS).

Example: Probit Estimates of The Probability of Holding Interest-Bearing Assets

Dependent Variable: HAVRASST
Method: ML - Binary Probit (Quadratic hill climbing)
Sample (adjusted): 3 3143
Included observations: 2842 after adjustments
Convergence achieved after 5 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	−5.569306	0.205488	−27.10289	0.0000
LNFINAST	0.638343	0.023260	27.44422	0.0000
PENSION	0.243126	0.069047	3.521156	0.0004
IRAS	0.208079	0.075085	2.771245	0.0056

Mean dependent var	0.620690	S.D. dependent var	0.485301
S.E. of regression	0.312913	Akaike info criterion	0.623138
Sum squared resid	277.8818	Schwarz criterion	0.631516
Log likelihood	−881.4797	Hannan-Quinn criter.	0.626160
Restr. log likelihood	−1886.308	Avg. log likelihood	−0.310162
LR statistic (3 df)	2009.656	McFadden R-squared	0.532696
Probability(LR stat)	0.000000		
Obs with Dep = 0	1078	Total obs	2842
Obs with Dep = 1	1764		

Example: Analysis – 1

- We can directly see that all three explanators are statistically significant (using the z-statistics).
- Also, all three explanators have positive coefficients. Increasing total financial assets, having a pension, and having an IRA all increase the probability of holding interest-bearing assets.

Example: Analysis – 2

- To assess the magnitude of the coefficient on PENSION, we need to conduct a follow-up calculation.
- A reasonable calculation would be to predict Prob(Y) when PENSION = 0 and when PENSION = 1, holding the other explanators fixed at their sample means.

DERIVING PROBIT/LOGIT

Deriving Probit/Logit

- Where do the Logit and Probit estimators come from?
- How does the latent variable Z determine whether Y = 1 or Y = 0?
- What role do the ε_i's play?

Assumptions for Y

We assume Y_i acts "as if" determined by latent variable Z.

$$Z_i = \beta_0 + \beta_1 X_{1i} + .. + \beta_K X_{Ki} + \varepsilon_i$$

$$Y_i = 1 \text{ if } Z_i > 0$$

$$Y_i = 0 \text{ if } Z_i \leq 0$$

But…

- Note: the assumption that the breakpoint falls at 0 is arbitrary.
- β_0 can adjust for whichever breakpoint you might choose to set.

Assumptions for residuals

- We assume we know the distribution of ε_i.
- In the probit model, we assume ε_i is distributed by the standard normal.
- In the logit model, we assume ε_i is distributed by the logistic.

Shocks – 1

- The key to Probit/Logit: since we know the distribution of ε_i , we can calculate the probability that a given observation receives a shock ε_i that pushes Z into the $Z > 0$ or $Z < 0$ region.

Shocks – 2

- Calculate $E(Z_i) = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_K X_{Ki}$
- Determine the regions of ε_i such that

$$E(Z_i) + \varepsilon_i < 0 \text{ or } E(Z_i) + \varepsilon_i > 0$$

- Using the distribution of ε_i , calculate the probability of drawing an ε_i from each region.

Example – 1

Suppose $E(Z_i) = 1$

If $\varepsilon_i >$ -1, then $E(Z_i) + \varepsilon_i > 0$

If $E(Z_i) + \varepsilon_i > 0$, then $Y_i = 1$

For the standard normal distribution,

what is the $Prob(\varepsilon_i >$ -1)?

Example – 2

- For the standard normal distribution, Prob(ε_i > -1) ≈ 0.83
- If $Z_i = 1$, we predict there is an 83% chance that Y = 1.

Example – 3

- For another example, suppose we are estimating a probit and $E(Z_i) = -2$. For what values of ε_i will $Z_i > 0$ (so $Y = 1$)?
- If $\varepsilon_i > 2$, $Z_i > 0$ (so $Y = 1$).
- For the standard normal distribution, $\text{Prob}(\varepsilon_i) > 2 \approx 0.025$. We predict a 2.5% chance that $Y = 1$.

General solution – 1

More generally, suppose ε_i has a cumulative density function F

That is, $Prob(\varepsilon_i < a) = F(a)$

$Prob(\varepsilon_i > a) = 1 - F(a)$

If F is symmetric, $1 - F(a) = F(-a)$

General solution – 2

More generally, $Prob(Y_i = 1)$

$$= Prob(\varepsilon_i > -E(Z_i))$$

$$= Prob(\varepsilon_i > -\hat{\beta}_0 - \hat{\beta}_1 X_{1i} - ..\hat{\beta}_K X_{Ki})$$

$$= 1 - Prob(\varepsilon_i < -\hat{\beta}_0 - \hat{\beta}_1 X_{1i} - ..\hat{\beta}_K X_{Ki})$$

$$= 1 - F(-\hat{\beta}_0 - \hat{\beta}_1 X_{1i} - ..\hat{\beta}_K X_{Ki})$$

$$= F(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + ..\hat{\beta}_K X_{Ki})$$

(for a symmetric distribution)

REVIEW

Our wish

- Frequently econometricians wish to estimate the probability that a discrete event occurs.
- The Linear Probability Model: estimating a probability by using a
linear model (e.g. OLS) with a dummy variable for Y.

Problems

Problems with the Linear Probability Model:

- OLS disturbances are heteroskedastic.
- OLS predictions range from - ∞ to + ∞. A probability needs to range from 0 to 1.

Solution

- Probit or Logit
- Assume a latent variable, Z, mediates between the explanators and the dummy variable Y.
- The higher Z is, the higher the probability that Y = 1.

Model

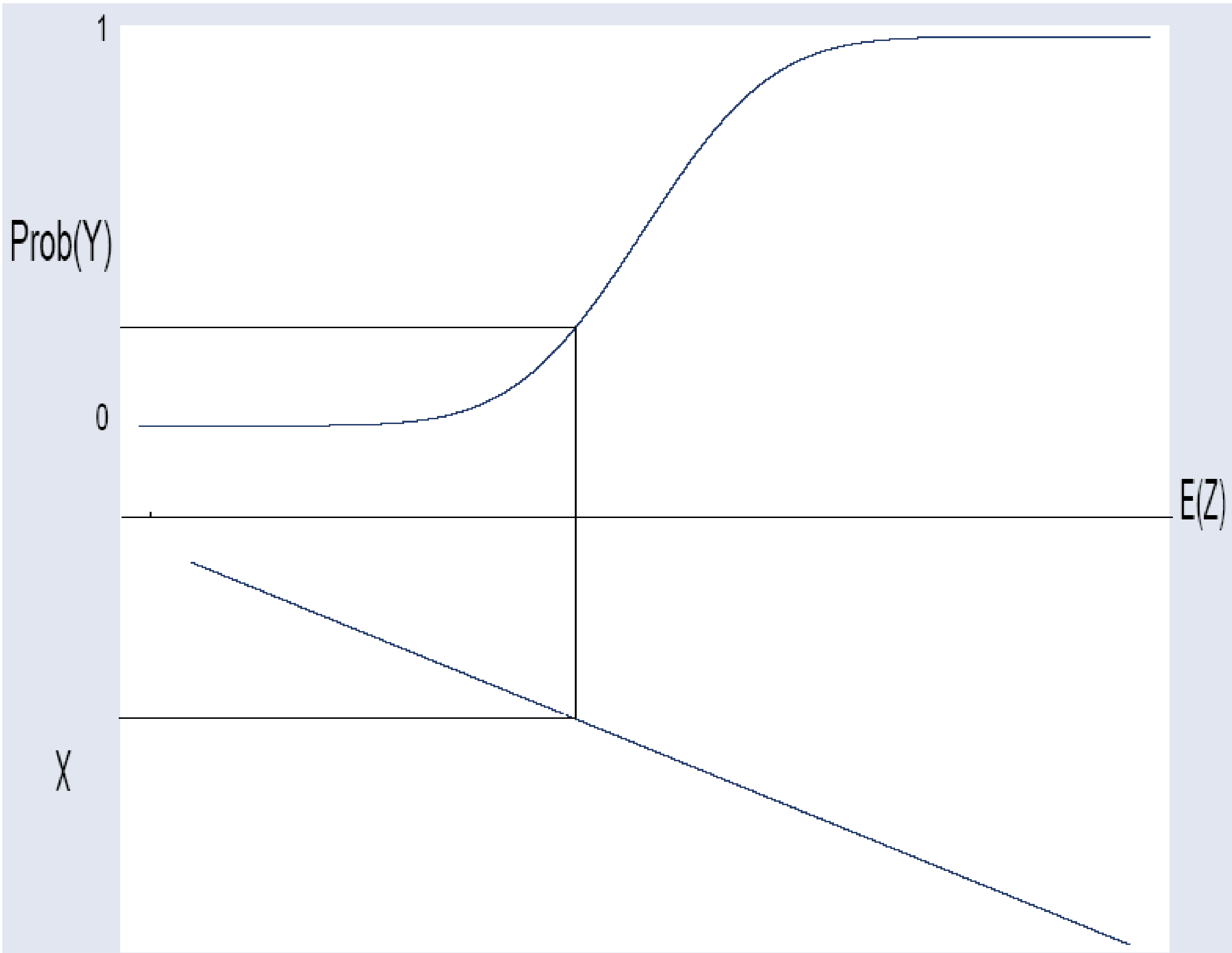
- To predict the Prob(Y) for a given X value, begin by calculating the fitted Z value from the predicted linear coefficients, for example:for only one explanator X :

$$E(Z) = \hat{Z}_i = \beta_0 + \beta_1 \hat{X}_i$$

- Then use the nonlinear function to translate the fitted Z value into a Prob(Y):

$$Prob(Y) = F(\hat{Z})$$

Prediction graph



Problems in Interpreting Magnitude

- The estimated coefficient relates X to Z. We care about the relationship between X and Prob(Y = 1).
- The effect of X on Prob(Y = 1) varies depending on Z.

Latent variable

We assume Y_i acts "as if" determined by latent variable Z.

$$Z_i = \beta_0 + \beta_1 X_{1i} + .. + \beta_K X_{Ki} + \varepsilon_i$$

$Y_i = 1$ if $Z_i > 0$

$Y_i = 0$ if $Z_i \leq 0$

Assumptions

- We assume we know the distribution of ε_i.
- In the probit model, we assume ε_i is distributed by the standard normal.
- In the logit model, we assume ε_i is distributed by the logistic.

General solution

$$Prob(Y_i = 1) = F(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + ..\hat{\beta}_K X_{Ki})$$

(for a symmetric distribution)

QUESTIONS?

THANK YOU FOR YOUR ATTENTION!