Computer Exercise 3: Binary Dependent Variable Models

Exercise 1 *Spread betting (such as handicap betting in William Hill) is often seen in sports gambling. The file pntsprd.csv contains the data related to the Las Vegas (renowned for gambling) point spread for college basketball games in the USA.*

1. *A linear probability model to estimate the probability that the favored team wins is*

$$P(favwin = 1|spread) = \beta_0 + \beta_1 spread. \tag{1}$$

Explain why, if the spread incorporates all relevant information, we expect $\beta_0 = 0.5$.

2. *Suppose now we use the probit model*

$$P(favwin = 1|spread) = \Phi(\beta_0 + \beta_1 spread). \tag{2}$$

If the spread incorporates all relevant information, what shall we expect β_0 *to be?*

3. *Use the data to estimate the linear probability model in (1). Test* $H_0 : \beta_0 = 0.5$ *against a two-sided alternative. Use both the usual and heteroskedasticity-robust standard errors.*

4. *Is spread statistically significant? What is the estimated probability that the favored team wins when* $spread = 10$*?*

5. *Test* $H_0 : \beta_0 = 0$ *in the probit model.*

6. *Use the probit model to estimate the probability that the favored team wins when* $spread = 10$*. Compare this with the LPM estimate.*

7. *Add the variables* $favhome$, $fav25$, *and* $und25$ *to the probit model. Can you notice an improvement in the McFadden R-squared measure?*

8. *Test the joint significance of these variables using the likelihood ratio test (*$2\left[\ln L(\hat{\beta}_U) - \ln L(\hat{\beta}_R)\right]$*) and Wald test. How many degrees of freedom should be used for the chi-square distribution? Interpret this result and check whether the spread incorporates all observable information prior to a game.*