Empirical Exercise: Chapter 11

Believe it or not, workers used to be able to smoke inside office buildings. Smoking bans were introduced in several areas during the 1990s. Supporters of these bans argued that in addition to eliminating the externality of secondhand smoke, they would encourage smokers to quit by reducing their opportunities to smoke. In this assignment, you will estimate the effect of workplace smoking bans on smoking, using data on a sample of 10,000 U.S. indoor workers from 1991 to 1993 in the file **Smoking**. The data set contains information on whether individuals were or were not subject to a workplace smoking ban, whether the individuals smoked, and other individual characteristics. A detailed description is given in **Smoking_Description**, available on the website.

a. Estimate the probability of smoking for (i) all workers, (ii) workers affected by workplace smoking bans, and (iii) workers not affected by workplace smoking bans.
b. What is the difference in the probability of smoking between workers affected by a workplace smoking ban and workers not affected by a workplace smoking ban? Use a linear probability model to determine whether this difference is statistically significant.
c. Estimate a linear probability model with *smoker* as the dependent variable and the following regressors: *smkban*, *female*, *age*, age^2, *hsdrop*, *hsgrad*, *colsome*, *colgrad*, *black*, and *hispanic*. Compare the estimated effect of a smoking ban from this regression with your answer from (b). Suggest an explanation, based on the substance of this regression, for the change in the estimated effect of a smoking ban between (b) and (c).
d. Test the hypothesis that the coefficient on *smkban* is 0 in the population version of the regression in (c) against the alternative that it is nonzero, at the 5% significance level.
e. Test the hypothesis that the probability of smoking does not depend on the level of education in the regression in (c). Does the probability of smoking increase or decrease with the level of education?
f. Repeat (c)–(e) using a probit model.
g. Repeat (c)–(e) using a logit model.
h.
   i. Mr. A is white, non-Hispanic, 20 years old, and a high school dropout. Using the probit regression and assuming that Mr. A is not subject to a workplace smoking ban, calculate the probability that Mr. A smokes. Carry out the calculation again, assuming that he is subject to a workplace smoking ban. What is the effect of the smoking ban on the probability of smoking?
   ii. Repeat (i) for Ms. B, a female, black, 40-year-old college graduate.
   iii. Repeat (i)–(ii) using the linear probability model.
   iv. Repeat (i)–(ii) using the logit model.
   v. Based on your answers to (i)–(iv), do the logit, probit, and linear probability models differ? If they do, which results make most sense? Are the estimated effects large in a real-world sense?