

Advanced Topics in Statistical Methods

PLSC 504

Exercise One

August 30, 2023

Part I: Logit vs. Probit Cage Match

Our first exercise is about binary response models. Despite your instructor's strident pleas, political scientists analyzing binary outcomes still split relatively evenly between logistic regression (logit) and probit. In Part I of this exercise, your general goal is to try to figure out – via simulation – if and under what conditions quantities of interest from logit and probit models diverge.

To keep things simple,¹ we'll focus on a regression model with a single binary predictor:

$$\begin{aligned}\Pr(Y_i = 1) &= f[\beta_0 + \beta_1 D_i] \\ Y_i &\in \{0, 1\} \\ D_i &\in \{0, 1\}\end{aligned}$$

We'll also stick to the case where the true data-generating process conforms to logistic regression; that is, where:

$$\Pr(Y_i = 1) = \frac{\exp(\beta_0 + \beta_1 D_i)}{1 + \exp(\beta_0 + \beta_1 D_i)} \quad (1)$$

Because the estimated β s for logit and probit have different scales, we'll concentrate here on *predicted probabilities*. The general plan for analysis is:

1. Simulate N observations of data that conform to Equation 1, above,
2. Fit a logit to those data, and generate predicted probabilities,
3. Fit a probit to the same data, and generate predicted probabilities,
4. Summarize the differences between the predicted probabilities from the two models (e.g., by looking at the mean of the N differences between them)
5. Repeat steps 1-4 many times, each time saving the summary in step 4,
6. Repeat steps 1-5, varying β_0 and β_1 (say, perhaps, over a range of values $\{-3, -2, \dots, 2, 3\}$) and N , and then
7. Collect all your findings, present them (graphically, in table(s), or otherwise) and discuss what you learned.

¹Because this is your first assignment, I'm going to be a bit more specific than I usually would about how to go about the simulation exercise.

Part II: “A covert ethnic-pride celebration for red-state whites of Northern European descent.”

In a 2019 publicity stunt, a Fox News commentator began her show by [attempting to drink a light-bulb-bedazzled steak through a straw](#). In her commentary on the bit, she noted that “it (the steak) has everything the Democrats hate. If I could have put an SUV on this, I would have.”

What does this have to do with political science? The [answer](#) ought to be [obvious](#).

Back in 2005, ABC News and the Washington Post commissioned a [poll](#) about public opinion on traffic. Among other things, pollsters asked 1204 lucky, randomly-selected Americans:

“What kind of vehicle do you usually drive – a car, an SUV, a pickup truck, or what?”

In this part, we’ll explore the political dynamics of car ownership, using the data from the 2005 ABC/WP poll. The main variable of interest is `cartype`, coded one for cars, two for SUVs, and three for pickup trucks. Covariates include dummy variables for urban residence, being married, having kids, and being `black` and/or `female`, as well as a naturally coded variable for `age` and an ordinal variable for level of `education`. Best of all, we also have two dichotomous variables for political party (`democrat` and `GOP`, with independents as our baseline) and a four-point ordinal scale indicating each respondent’s approval or disapproval for then-President Bush. All data are available on the PLSC 504 [Github repository](#), in the folder creatively names “Exercises.”

Exercise

1. After examining summary statistics, estimate a multinomial logit (MNL) model of vehicle type ownership. Report your estimation results, and interpret these findings, in statistical and substantive terms. Are the results in the “expected directions”? Discuss their statistical significance.
2. Using the MNL results, generate and examine the predicted probabilities of each type of vehicle, across the range of values for *one* of your more important independent variables, using tables or graphs of the probabilities. Interpret these results in substantive terms.
3. Examine / test for whether the data/model conform to the MNL model’s IIA assumption. Use whatever methods you can / are aware of, and discuss your findings on this point in both statistical and substantive terms.
4. Irrespective of the results of the IIA test(s), reestimate the same specification using HEV and multinomial probit models. Compare those findings to the MNL results, and briefly discuss similarities and differences. You might find the [vignettes](#) for the `mlogit` package useful here.
5. Finally, briefly discuss *in substantive terms* what your statistical conclusions suggest for the relationship between political ideology/preferences and automobile ownership.

This assignment is due *electronically*, as a *PDF file*, at 11:59 p.m. ET on Thursday, September 7, 2023; you can submit your homework by emailing copies **both** to Dr. Zorn (zorn@psu.edu) and Mr. Morse (nam@psu.edu). This assignment is worth 50 possible points.