POLS/CSSS 510:
Maximum Likelihood Methods
for the Social Sciences

Problem Set 4

Prof. Christopher Adolph, Political Science and CSSS

Fall Quarter 2024

Due in class on Wednesday 20 November 2024

General instructions for homeworks: For any exercises done with R or other statistical packages, you should attach your code at the end of the document. All other materials should be collated in order by problem. The most readable and elegant format for homework answers incorporates student comments and graphics into a seamless narrative, as one would see in an article or textbook.

In this assignment you will use data from the 1992 American National Election Study. The data are taken from a nationally representative sample of the US population. The survey was conducted in the fall of 1992 with a reinterview of the respondents following the election. In the data you will be using, only the vote report is taken from the post-election interview.

The dataset nes92.csv contains information on the vote for President and approval of then-President Bush along with several variables selected to capture the range of influences that current work on voting behavior expects to be important. Descriptions of the variables are at the end of this assignment.

Problem 1: Modeling presidential approval with ordered probit

a. [5 points] Use an ordered probit model to explain respondents' approval of Bush. Using the `llk.oprobit()` function provided in class and `optim()`, fit the model to the variable `bushapp` with `milforce`, `rbdist`, `econ`, `partyid`, and `yrsofed` as the only covariates. Report the estimated parameters, their standard errors, and the value of the log likelihood at its maximum.

b. [15 points] Calculate the probabilities that a respondent who identified themselves as a strong Democrat (`partyid` $= -3$) said they "Strongly Disapproved," "Disapproved," "Approved," or "Strongly Approved" of President Bush given their support or opposition to military force (`milforce` $= \{1, 2, 3, 4, 5\}$), holding the remaining covariates at their mean values. Calculate the probabilities again over the same range of responses for milforce for a respondent who identified themselves as a strong Republican (`partyid` $= +3$). You should have two sets of 20 probabilities, or 40 probabilities total.

c. [15 points] Present the expected probabilities for each category you calculated in part **b** in a way that would not require your reader to know anything about ordered probit models. Be sure to present confidence intervals, in a fashion of your choosing.

d. [10 points] In this part, you are asked to summarize the model estimated in part a. using just two simulated first differences. You may not alter the model itself to eliminate or collapse categories, as this would throw away useful information. ***i.*** Simulate the increase in probability of *either* approving or strongly approving of Bush given a shift from strong opposition to the use of force to strong support for a respondent who is a strong Democrat and otherwise average. Provide a 95% confidence interval. ***ii.*** Simulate the increase in probability of *either* approving or strongly approving of Bush given a shift from strong opposition to the use of force to strong support for a respondent who is a strong Republican and otherwise average. Provide a 95% confidence interval.

e. [5 points] Based upon your work above, briefly discuss of the results of the model.

f. [10 points] The `llk.oprobit()` function is written for ordered data with four categories. Alter the function so that it can be used to estimate a three category

ordered probit. Using R, create a new variable econ3 which collapses econ into a three-category variable.[1] Fit a model to the variable econ3 with partyid as the only covariates to see the effect of an individual's party id on their perception of the economy. Simply run the model and report the estimated parameters. Perform this fit using your altered function and optim(). ***Bonus [+10 points]:*** Modify the likelihood function so that it works generically for any number of categories.

g. **[10 points]** *After* you have completed the above parts of the problem, look carefully at the example given in the help for oprobitsimev in the simcf package. Re-estimate your model using the polr function from the MASS library. polr uses an alternative parameterization of the ordered probit which drops the model's constant in favor of an additional cutpoint (i.e., instead of using a constant and $k - 2$ cutpoints for a k category response variable, polr uses $k - 1$ cutpoints and no constant). Note the changes in how we use oprobitsimev when using polr to estimate ordered probit. Then show that whether we use optim or polr, we end up with exactly the same predicted probabilities for each category of our response for any given counterfactual scenario.

---

[1] Specifically, map the following values from econ to econ3: $\{1, 2\} \to \{1\}$; $\{3\} \to \{2\}$; and $\{4, 5\} \to \{3\}$.

Variable	Description
vote92	Vote 1992: 0 = Bush, 1 = Clinton, 2 = Perot
bushapp	Bush Approval, 1992: 0 = Strongly Disapprove, 1 = Disapprove, 2 = Approve, 3 = Strongly Approve
rlibcon	Ideological self-placement: 1 = very liberal, ... 7 = very conservative
rbdist	Ideological distance between respondent and Bush in absolute value
econ	Economy worse than year ago? 1 = much better, 2 = better, 3 = same, 4 = worse, 5 = much worse
milforce	Opposition to military force: 1 = would use force, ... 5 = would not use force
gulfwar	Was the Gulf War worth the cost? 0 = not worth the cost, 1 = worth cost
partyid	−3 = Strong Democrat, −2 = Democrat, −1 = Leans Democratic, 0 = Independent, ... 3 = Strong Republican
yrsofed	Years of education in actual years: 0–17
nonwhite	0 = white, 1 = nonwhite
econ3	Recoding of econ into three categories: 1 = better, 2 = same, 3 = worse. *Construction of this variable left to the student.*