Problem Set 3 - Laura McPhillips

Applied Stats II

Due: March 28, 2022

Question 1

We are interested in how governments' management of public resources impacts economic prosperity. Our data come from Alvarez, Cheibub, Limongi, and Przeworski (1996) and is labelled `gdpChange.csv` on GitHub. The dataset covers 135 countries observed between 1950 or the year of independence or the first year forwhich data on economic growth are available ("entry year"), and 1990 or the last year for which data on economic growth are available ("exit year"). The unit of analysis is a particular country during a particular year, for a total $>$ 3,500 observations.

- Response variable:
  - `GDPWdiff`: Difference in GDP between year t and $t-1$. Possible categories include: "positive", "negative", or "no change"
- Explanatory variables:
  - `REG`: 1=Democracy; 0=Non-Democracy
  - `OIL`: 1=if the average ratio of fuel exports to total exports in 1984-86 exceeded 50%; 0= otherwise

Import data and libraries

```
data <- read.csv("gdpChange.csv", stringsAsFactors = F)

library(nnet)
library(MASS)
library(stargazer)
```

Please answer the following questions:

1. Construct and interpret an unordered multinomial logit with `GDPWdiff` as the output and "no change" as the reference category, including the estimated cutoff points and coefficients.

```
data$GDPWdiff <- replace(data$GDPWdiff, data$GDPWdiff < 0, -1)
data$GDPWdiff <- replace(data$GDPWdiff, data$GDPWdiff > 0, 1)

data$GDPWdiff <- factor((data$GDPWdiff),
                        levels = c("-1", "0", "1"),
                        labels = c("negative", "no change", "positive"))


data$GDPWdiff <- relevel(data$GDPWdiff, ref = "no change")

mult.log <- multinom(GDPWdiff ~ REG + OIL, data = data)
summary(mult.log)

```

Table 1:

	Dependent variable:	
	negative	positive
	(1)	(2)
REG	1.379*	1.769**
	(0.769)	(0.767)
OIL	4.784	4.576
	(6.885)	(6.885)
Constant	3.805***	4.534***
	(0.271)	(0.269)
Akaike Inf. Crit.	4,690.770	4,690.770
Note:	*p<0.1; **p<0.05; ***p<0.01	

Changing from non-democracy to democracy, the log odds for countries with negative GDP growth, relative to countries with no GDP growth, would expect to increase by 1.37, holding all other variables constant.

Changing from non-democracy to democracy, the log odds for countries with positive GDP growth, relative to countries with no GDP growth, would expect to increase by 1.76, holding all other variables constant.

If fuel exports were to exceed more than 50% of the total exports ratio, the log odds for countries with negative GDP growth, relative to countries with no GDP growth, would expect to increase by 4.78 holding all other variables constant. (not significant)

If fuel exports were to exceed more than 50% of the total exports ratio, the log odds for countries with positive GDP growth, relative to countries with no GDP growth, would expect increase by 4.5 holding all other variables constant. (not significant)

(For the negative category, with REG and OIL at 0, log odds of GDP growth is 3.8. For the positive category, with REG and OIL at 0, log odds of GDP growth is 4.5.)

2. Construct and interpret an ordered multinomial logit with `GDPWdiff` as the outcome variable, including the estimated cutoff points and coefficients.

```
data <- read.csv("gdpChange.csv", stringsAsFactors = F)
data$GDPWdiff <- replace(data$GDPWdiff, data$GDPWdiff < 0, -1)
data$GDPWdiff <- replace(data$GDPWdiff, data$GDPWdiff > 0, 1)

data$GDPWdiff <- factor((data$GDPWdiff),
                        levels = c("-1", "0", "1"),
                        labels = c("negative", "no change", "positive"))
ord.log <- polr(GDPWdiff ~ REG + OIL, data = data, Hess = TRUE)
summary(ord.log)
```

Table 2:

	Dependent variable:
	GDPWdiff
REG	0.398***
	(0.075)
OIL	−0.199*
	(0.116)
Observations	3,721
Note:	*p<0.1; **p<0.05; ***p<0.01

Changing from non-democracy to democracy, increased the log odds of GDP growth by 0.39.

Exceeding more than 50% of fuel exports to total exports ratio, decreased the log odds of GDP growth by 0.19.

Question 2

Consider the data set `MexicoMuniData.csv`, which includes municipal-level information from Mexico. The outcome of interest is the number of times the winning PAN presidential candidate in 2006 (`PAN.visits.06`) visited a district leading up to the 2009 federal elections, which is a count. Our main predictor of interest is whether the district was highly contested, or whether it was not (the PAN or their opponents have electoral security) in the previous federal elections during 2000 (`competitive.district`), which is binary (1=close/swing district, 0="safe seat"). We also include `marginality.06` (a measure of poverty) and `PAN.governor.06` (a dummy for whether the state has a PAN-affiliated governor) as additional control variables.

(a) Run a Poisson regression because the outcome is a count variable. Is there evidence that PAN presidential candidates visit swing districts more? Provide a test statistic and p-value.

```
dataset <- read.csv("MexicoMuniData.csv")
mexico_poisson <- glm(PAN.visits.06 ~ competitive.district + marginality
    .06 + PAN.governor.06, family = "poisson", data = dataset)
summary(mexico_poisson)
```

Changing from a safe district to a swing district decreases the log count of a PAN presidential candidate visit by 0.081, but this is not significant.
Test statistic = -0.47 P-value = 0.63
The p-value for competitive.district is not statistically significant, therefore there is not enough evidence to say that PAN presidential candidates visit swing districts more.

Table 3:

	Dependent variable:
	PAN.visits.06
competitive.district	−0.081
	(0.171)
marginality.06	−2.080***
	(0.117)
PAN.governor.06	−0.312*
	(0.167)
Constant	−3.810***
	(0.222)
Observations	2,407
Log Likelihood	−645.606
Akaike Inf. Crit.	1,299.213
Note:	*p<0.1; **p<0.05; ***p<0.01

(b) Interpret the `marginality.06` and `PAN.governor.06` coefficients.

```
cfs <- exp(coef(mexico_poisson))
cfs
```

Table 4:

(Intercept)	competitive.district	marginality.06	PAN.governor.06
0.022	0.922	0.125	0.732

A one unit increase in poverty decreases PAN presidential candidate visits by a multiplicative factor of 0.12 , when all other variables are held constant.

Changing from no governor to PAN governor decreases PAN presidential candidate visits by a multiplicative factor of 0.73, holding all other variables constant.

(They are decreasing because $\beta 1$ and $\beta 2$ are negative.)

(c) Provide the estimated mean number of visits from the winning PAN presidential candidate for a hypothetical district that was competitive (`competitive.district`=1), had an average poverty level (`marginality.06` = 0), and a PAN governor (`PAN.governor.06`=1).

```
pred <- data.frame(competitive.district = 1,
                   marginality.06 = 0,
                   PAN.governor.06 = 1)

predict(mexico_poisson, newdata = pred, type = "response")
```

Mean number of visits for winning PAN presidential candidate for the hypothetical district above = 0.0149.