Problem Set 3 - My Answer

Junjie LIU[1,*]

23 March, 2024

[1] Department of Political Science, Trinity College Dublin, 2 Clare, Street, Dublin 2, Ireland

[*] Correspondence: Junjie LIU <liuj13@tcd.ie>

Instructions

- Please show your work! You may lose points by simply writing in the answer. If the problem requires you to execute commands in `R`, please include the code you used to get your answers. Please also include the `.R` file that contains your code. If you are not sure if work needs to be shown for a particular problem, please ask.
- Your homework should be mitted electronically on GitHub in `.pdf` form.
- This problem set is due before 23:59 on Sunday March 24, 2024. No late assignments will be accepted.

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

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.

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

```r
# Question 1
library(nnet)
library(MASS)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following object is masked from 'package:MASS':
##
##     select

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
gdpChange <- read.csv("../../../datasets/gdpChange.csv")

gdpChange$GDPWdiff <- ifelse(gdpChange$GDPWdiff < 0, "negative",
                      ifelse(gdpChange$GDPWdiff == 0, "no_change", "positive"))
gdpChange <- gdpChange %>% mutate(OIL=as.factor(OIL), REG=as.factor(REG),
                    COUNTRY=as.factor(COUNTRY), GDPWdiff=as.factor(GDPWdiff))
gdpChange$GDPWdiff <- relevel(gdpChange$GDPWdiff, ref="no_change")

model_1 <- multinom(GDPWdiff ~ REG + OIL, data=gdpChange)
```

```
## # weights:  12 (6 variable)
## initial  value 4087.936326
## iter  10 value 2340.076844
## final  value 2339.385155
## converged
```

```r
summary(model_1)
```

```
## Call:
## multinom(formula = GDPWdiff ~ REG + OIL, data = gdpChange)
##
```

```
## Coefficients:
##          (Intercept)     REG1     OIL1
## negative    3.805370 1.379282 4.783968
## positive    4.533759 1.769007 4.576321
##
## Std. Errors:
##          (Intercept)      REG1     OIL1
## negative   0.2706832 0.7686958 6.885366
## positive   0.2692006 0.7670366 6.885097
##
## Residual Deviance: 4678.77
## AIC: 4690.77
```

```r
model_2 <- polr(GDPWdiff ~ REG + OIL, data=gdpChange)
summary(model_2)
```

```
##
## Re-fitting to get Hessian

## Call:
## polr(formula = GDPWdiff ~ REG + OIL, data = gdpChange)
##
## Coefficients:
##        Value Std. Error t value
## REG1  0.4102    0.07518   5.456
## OIL1 -0.1788    0.11546  -1.549
##
## Intercepts:
##                      Value    Std. Error t value
## no_change|negative  -5.3199   0.2523   -21.0878
## negative|positive   -0.7036   0.0476   -14.7933
##
## Residual Deviance: 4686.606
## AIC: 4694.606
```

Based on the model 1, we can find that

1. For the negative, the log-odds is denoted by: $y = 3.81 + 1.38 \times \text{REG1} + 4.78 \times \text{OIL1}$
2. For the positive, the log-odds is denoted by: $y = 4.53 + 1.77 \times \text{REG1} + 4.58 \times \text{OIL1}$

For the ordered model, we can find that

1. The "no change|negative" at -5.32 with a t-value of -21.09 would be the threshold between "no change" and "negative" categories
2. The "negative|positive" at -0.70 with a t-value of -14.79 is the threshold between "negative" and "positive" categories

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.

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

(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).

```r
mexico <- read.csv("../../../datasets/MexicoMuniData.csv")
mexico <- mexico %>% mutate(PAN.governor.06=as.factor(PAN.governor.06),
                            competitive.district=as.factor(competitive.district))

model_3 <- glm(PAN.visits.06 ~ PAN.governor.06 + competitive.district + marginality.06,
               data = mexico, family = poisson(link = "log"))
summary(model_3)
```

```
##
## Call:
## glm(formula = PAN.visits.06 ~ PAN.governor.06 + competitive.district +
##     marginality.06, family = poisson(link = "log"), data = mexico)
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -3.81023    0.22209 -17.156   <2e-16 ***
## PAN.governor.061      -0.31158    0.16673  -1.869   0.0617 .
## competitive.district1 -0.08135    0.17069  -0.477   0.6336
## marginality.06        -2.08014    0.11734 -17.728   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1473.87  on 2406  degrees of freedom
## Residual deviance:  991.25  on 2403  degrees of freedom
## AIC: 1299.2
##
## Number of Fisher Scoring iterations: 7
```

```r
coeff <- coefficients(model_3)
output <- exp(sum(coeff * c(1, 1, 1, 0)))
print(paste("The estimated mean number of visits is", output))
```

```
## [1] "The estimated mean number of visits is 0.0149481810354763"
```

Based on the model 3, we can find that

1. The t-value and p-value for "competitive district 1" is -0.48 and 0.63, where the the p-value is much greater than 0.05. There is not statistically significant
2. The coefficient for "marginality 06" is -2.08, which is statistically significant. This negative coefficient indicates that as marginality increases by one unit, the log count of visits from the PAN presidential candidates is expected to decrease by 2.08
3. The coefficient for "PAN governor 06" is -0.31 with a p-value of 0.6317, which is not significant
4. The estimated mean number of visits is 0.015