

# Problem Set 3

Daniel Murray (13303981)

Due: March 26, 2023

## 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 submitted electronically on GitHub in .pdf form.
- This problem set is due before 23:59 on Sunday March 26, 2023. 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 for which 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.

**Answer:**

The data was read in, inspected and wrangled to change the explanatory variables `REG` and `OIL` to factors and the outcome variable `GDPWdiff` to an unordered factor with three levels: "negative", "no change", and "positive". The category "no change" was then set as the reference category. See code used below:

```
1 # Read in data
2 data <- read.csv("data/gdpChange.csv")
3
4 # Inspect data
5 head(data)
6 summary(data)
7 str(data)
8
9 # Wrangle data
10 data$REG <- as.factor(data$REG)
11 data$OIL <- as.factor(data$OIL)
12 data$GDPWdiff2[data$GDPWdiff < 0] <- "negative"
13 data$GDPWdiff2[data$GDPWdiff == 0] <- "no change"
14 data$GDPWdiff2[data$GDPWdiff > 0] <- "positive"
15 data$GDPWdiff <- data$GDPWdiff2
16 data$GDPWdiff2 <- NULL
17
18 # Change GDPWdiff to unordered factor with "no change" as reference
19 data$GDPWdiff <- factor(data$GDPWdiff, ordered = FALSE)
20 data$GDPWdiff <- relevel(data$GDPWdiff, ref="no change")
```

Following this, a multinomial logit regression was run using the below code:

```
1 # Run unordered multinomial logit regression model
2 mod_unordered <- multinom(GDPWdiff ~ REG + OIL, data = data)
3 summary(mod_unordered)
4 stargazer(mod_unordered, type="latex")
```

This produced the output in Table 1 below. The output shows that being a democracy leads to a an increase of 1.379 in the log odds of moving from "no change" to "negative", on average, holding the other variables constant. Having fuel exports make up over

50% of total exports on average during 1984-1986 increases the log odds of moving from "no change" to "negative" by 4.787 on average, holding other variables constant. For a non-democracy with less than 50% of average total exports coming from fuel during 1984-1986, the log odds of moving from "no change" to "negative" is 3.805 on average.

Similarly, being a democracy leads to an increase of 1.769 in the log odds of moving from "no change" to "positive", on average, holding the other variables constant. Having fuel exports make up over 50% of total exports on average during 1984-1986 increases the log odds of moving from "no change" to "positive" by 4.576 on average, holding other variables constant. For a non-democracy with less than 50% of average total exports coming from fuel during 1984-1986, the log odds of moving from "no change" to "positive" is 4.534 on average.

Table 1:

	<i>Dependent variable:</i>	
	negative	positive
	(1)	(2)
REG1	1.379* (0.769)	1.769** (0.767)
OIL1	4.784 (6.885)	4.576 (6.885)
Constant	3.805*** (0.271)	4.534*** (0.269)
Akaike Inf. Crit.	4,690.770	4,690.770
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

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

### **Answer:**

To run an ordered multinomial logit regression, the outcome variable was first changed to an ordered factor with levels set in the following order: "negative", "no change", "positive". The below code was used:

```

1 # Change GDPWdiff to ordered factor
2 data$GDPWdiff <- factor(data$GDPWdiff, ordered = TRUE)
3 data$GDPWdiff <- factor(data$GDPWdiff,
4                           levels = c("negative", "no change", "positive"),
5                           labels = c("Negative", "No Change", "Positive"))

```

Next, the model was run using the below code:

```

1 # Run ordered multinomial logit regression model
2 mod_ordered <- polr(GDPWdiff ~ REG + OIL, data = data, Hess = TRUE)
3 summary(mod_ordered)

```

This produced the output in Figure 1 below. The output shows that being a democracy is associated with a 0.3985 increase in the log odds of moving from "negative" to "no change", or from "no change" to positive", holding the other variable `OIL` constant. Having fuel exports make up over 50% of total exports on average during 1984-1986 is associated with a 0.1987 decrease in the log odds of moving from "negative" to "no change", or from "no change" to positive", holding the other variable `REG` constant. The cutoff where our estimated outcome changes from "negative" to "no change" is where the value of log odds is equal to -0.7312. The cutoff where our estimated outcome changes from "no change" to "positive" is where the value of log odds is equal to -0.7105.

Figure 1: Output of ordered multinomial regression model

```

Call:
polr(formula = GDPWdiff ~ REG + OIL, data = data, Hess = TRUE)

Coefficients:
              Value Std. Error t value
REG1  0.3985      0.07518   5.300
OIL1 -0.1987      0.11572  -1.717

Intercepts:
              Value      Std. Error t value
Negative|No Change -0.7312      0.0476  -15.3597
No Change|Positive -0.7105      0.0475  -14.9554

Residual Deviance: 4687.689
AIC: 4695.689

```

## 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.

**Answer:**

A Poisson regression was run using the below code, producing the output shown in Figure 2.

```
1 # Run poisson regression
2 vis_poisson <- glm(PAN.visits.06 ~ competitive.district + marginality.06
  + PAN.governor.06,
3                       data=data2, family=poisson)
4 summary(vis_poisson)
```

Figure 2: Output of Poisson regression model

```
Call:
glm(formula = PAN.visits.06 ~ competitive.district + marginality.06 +
    PAN.governor.06, family = poisson, data = data2)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2309  -0.3748  -0.1804  -0.0804   15.2669

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -3.81023    0.22209  -17.156  <2e-16 ***
competitive.district1 -0.08135    0.17069   -0.477    0.6336
marginality.06   -2.08014    0.11734  -17.728  <2e-16 ***
PAN.governor.061  -0.31158    0.16673   -1.869    0.0617 .
---
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
```

The z-score test statistic (-0.477) indicates that there is not enough evidence to suggest that PAN presidential candidates visit swing districts more. This test score results in a P-value of 0.6336, which is above our critical threshold for rejecting the null hypothesis that being a competitive district has no effect on the number of visits.

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

**Answer:**

Exponentiating the coefficients generated by our model gives us the values in Figure 3. A one unit increase in the marginality of a municipality changes the average number of visits by a multiplicative factor of 0.125, holding the other variables `PAN.governor.06` and `competitive.district` constant. A one unit increase in `PAN.governor.06` (i.e. having a PAN governor) changes the average number of visits by a multiplicative factor of 0.732, holding the other variables `marginality.06` and `competitive.district` constant.

Figure 3: Exponentiated coefficients

(Intercept)	<code>competitive.district1</code>	<code>marginality.06</code>	<code>PAN.governor.061</code>
0.02214298	0.92186932	0.12491227	0.73228985

- (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`).

**Answer:**

The estimated mean number of visits from the winning PAN presidential candidate for a hypothetical district that was competitive, had an average poverty level and a PAN governor was calculated using the below code:

```
1 # Extract coefficients
2 cfs <- coef(vis_poisson)
3 cfs

1 # Predicted no. of visits for a competitive district that is zero-
  marginal and has a PAN governor
2 exp(cfs[1] + cfs[2]*1 + cfs[3]*0 + cfs[4]*1)
```

The value was calculated as 0.0149 visits.