Problem Set 3

Valeriia Babaian

March 24, 2024

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.

```r
# subsetting the DV values and transforming to a factor:
gdp_data$GDPWdiff <- as.factor(ifelse(gdp_data$GDPWdiff > 0, "positive",
                      ifelse(gdp_data$GDPWdiff < 0, "negative", "no change")))

ftable(gdp_data$GDPWdiff)

# subsetting the regime values and transforming to a factor:
gdp_data$REG <- as.factor(ifelse(gdp_data$REG == 0, "Non-Democracy", "Democracy"))
ftable(gdp_data$REG)

# subsetting the fuel exports share values and transforming to a factor:
gdp_data$OIL <- as.factor(ifelse(gdp_data$OIL > 0.5, "1", "0"))
ftable(gdp_data$OIL)

### 1: an unordered multinomial logit

# set a reference level for the DV
gdp_data$GDPWdiff <- relevel(gdp_data$GDPWdiff, ref = "no change")

# run model
mult.log <- multinom(GDPWdiff ~ REG + OIL, data = gdp_data)
summary(mult.log)
exp(coef(mult.log))
```

   According to the results provided in the Table 1, both growth and decrease in GDP are on average more modest in non-democracies, holding the share of fuel exports constant. Exponential values indicate that the odds of GDP decline are almost 4 times lower in non-democracies, and the odds of GDP growth in non-democracies are almost 6 times lower in non-democracies, all compared with the odds of the constant GDP.

   The average effect of fuel exports on the GDP dynamics is statistically indifferentiable from 0.

Table 1:

	Dependent variable:	
	negative	positive
	(1)	(2)
REGNon-Democracy	−1.339*	−1.728**
	(0.754)	(0.753)
OIL1	7.313	7.105
	(24.294)	(24.294)
Constant	5.141***	6.259***
	(0.704)	(0.703)
Akaike Inf. Crit.	4,690.730	4,690.730
Note:	*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.

```
gdp_data$GDPWdiff <- relevel(gdp_data$GDPWdiff, ref = "negative")
ord.log <- polr(GDPWdiff ~ REG + OIL, data = gdp_data, Hess = TRUE)
summary(ord.log)

ci <- confint(ord.log)

exp(cbind(OR = coef(ord.log), ci))
```

The ordered model output in the Tables 2 and 3 shows that on average, odds of GDP growth in non-democracies are lower than in democracies, with share of fuel exports hold constant (lesser by one third). With regime type hold constant, high fuel exports (more than half of total exports) are negatively associated with a positive GDP dynamic, as compared with lesser fuel exports, by roughly one fifth.

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,

Table 2:

	Dependent variable:
	GDPWdiff
REGNon-Democracy	-0.398^{***}
	(0.075)
OIL1	-0.199^{*}
	(0.116)
Observations	3,721
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3:

	OR	2.5 %	97.5 %
REGNon-Democracy	0.6713399	0.5790108	0.7775131
OIL1	0.8197839	0.6545844	1.0306562

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.

```
mod.ps <- glm(PAN.visits.06 ~ competitive.district +
                  marginality.06 + PAN.governor.06, data = mexico_elections
    , family = poisson)
stargazer::stargazer(mod.ps, out = "2.1.tex")

summary(mod.ps)
cfs <- exp(coef(mod.ps))
```

No, there is no evidence of more frequent visits to swing districts by winning PAN presidential candidates. The corresponding test statistic from the output generated in the regression summary is -0.477 (as compared with -0.081 coefficient estimate), and the p-value is 0.634. Thus, we cannot reject the null hypothesis that number of visits to a district is not related to it being swing.

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

Table 4:

	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

Poorer districs on average get a roughly 8 times lesser number of PAN presidential candidates, holding the competitiveness of the disctrict and the party membership of the governor constant. At 0.1 significance level, having a PAN governor decreases number of visits by their winning presidential candidates by roughly a quarter, with poverty and competitiveness of the district hold constant. The estimates are obtained from the exponential 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).

The prediction equation would have the following form:

$\hat{PAN.visits}.06 = e^{-3.81+(-0.081\times1)+(-2.08\times0)+(-3.81\times1)}$

```
exp(coef(mod.ps)[1] + coef(mod.ps)[2]*1 + coef(mod.ps)[3]*0 + coef(mod.ps
    )[4]*1)
# 0.01494818

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

```
8
9 # check with predict() function
10 predict(mod.ps, newdata = pred, type = "response")
11 #0.01494818
```

The estimated mean number of visits from the winning PAN presidential candidate for a hypothetical district with the abovementioned conditions is 0.015.