Problem Set 3 - Applied Stats II

Robert Baker

2024-03-21

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.

```
# Loading data
gdp_data <-
  read.csv("https://raw.githubusercontent.com/ASDS-TCD/StatsII_Spring2024/main/datasets/gdpChange.csv")

# Creating factor variables for outcome
gdp_data <- gdp_data %>%
  mutate(GDPcat = case_when(
    GDPWdiff == 0 ~ "no change",
    GDPWdiff > 0  ~ "positive",
    GDPWdiff < 0  ~ "negative"
  )) %>%
  mutate(GDPcat = factor(GDPcat, levels = c("no change", "positive", "negative")))

# Running Unordered Multinomial Logistic Regression
unordered_logit <- multinom(GDPcat ~ REG + OIL, data = gdp_data)
```

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

```
summary(unordered_logit)
```

```
## Call:
## multinom(formula = GDPcat ~ REG + OIL, data = gdp_data)
##
## Coefficients:
```

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

The above estimates show that an increase in REG by one unit is associated with an increase in the log odds of transitioning to a positive change in economic prosperity by 1.77. An increase in OIL by one unit increases the log odds by 4.58, with the intercept for this category at 4.53. This suggests that both democratic governance and higher reliance on fuel exports enhance the likelihood of positive economic growth compared to no change.

Similarly, for transitions to 'negative' growth, an increase in REG by one unit increases the log odds by 1.38, while an increase in OIL by one unit increases the log odds by 4.78, with the category's intercept at 3.81. This may indicate the same factors that contribute to positive growth also increase the likelihood of negative growth compared to no change.

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

```r
# Adjusting to specify an ordered factor
gdp_data <- gdp_data %>%
  mutate(GDPcat = case_when(
    GDPWdiff < 0 ~ "negative",
    GDPWdiff == 0 ~ "no change",
    GDPWdiff > 0 ~ "positive"
  )) %>%
  mutate(GDPcat = factor(GDPcat, levels = c("negative", "no change", "positive"),
                         ordered = TRUE))

# checking
str(gdp_data$GDPcat)
```

```
##  Ord.factor w/ 3 levels "negative"<"no change"<..: 3 1 3 3 3 3 1 3 1 3 ...
```

```r
# Running Ordered Multinomial Logistic Regression
ordered_logit <- polr(GDPcat ~ REG + OIL, data=gdp_data)
summary(ordered_logit)
```

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

```
## Call:
## polr(formula = GDPcat ~ REG + OIL, data = gdp_data)
##
## Coefficients:
##       Value Std. Error t value
## REG  0.3985    0.07518   5.300
## OIL -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
```

The model indicates that a one-unit increase in REG is associated with an increase in the log odds of transitioning from 'negative' to 'no change' and from 'no change' to 'positive' economic conditions by approximately 0.40 (std. error = 0.08, t = 5.30). Conversely, a one-unit increase in OIL decreases the log odds of these transitions by -0.20 (std. error = 0.12, t = -1.72).

The intercept for transitioning from 'negative' to 'no change' is at -0.73 (std. error = 0.05, t = -15.36), and from 'no change' to 'positive' is at -0.71 (std. error = 0.05, t = -14.96). These thresholds demarcate the latent variable continuum - showing the points at which the probability of being in one category shifts to the next. The model assumes an order among economic changes.

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 can- didate 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.

```r
# Loading data
mexico_elections <-
  read.csv("https://raw.githubusercontent.com/ASDS-TCD/StatsII_Spring2024/main/datasets/MexicoMuniData.

# Estimating Poisson Regression Model
mexico_poisson <- glm(PAN.visits.06 ~ competitive.district + marginality.06 + PAN.governor.06,
                      data = mexico_elections, family=poisson)
summary(mexico_poisson)
```

```
##
## Call:
## glm(formula = PAN.visits.06 ~ competitive.district + marginality.06 +
##     PAN.governor.06, family = poisson, data = mexico_elections)
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -3.81023    0.22209 -17.156   <2e-16 ***
## competitive.district -0.08135    0.17069  -0.477   0.6336
## marginality.06       -2.08014    0.11734 -17.728   <2e-16 ***
## PAN.governor.06      -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
```

We can see from the Poisson Regression there is no significant evidence that PAN presidential candidates visited swing districts more than safe districts. The test statistic for 'competitive.district' is z = -0.48, which has a p-value of 0.63, which is greater than a significance level of 0.05. This indicates the test statistic is not large enough to reject the null hypothesis at a 5% significance level.

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

Interpreting the 'marginality.06' coefficient, a one-unit increase in the marginality index, reflecting higher poverty, leads to a multiplicative decrease in the expected number of PAN candidate visits by approximately 0.12, when holding the effects of having a PAN governor and the competitiveness of the district constant. For the 'PAN.governor.06' coefficient, the presence of a PAN-affiliated governor is associated with an expected number of visits being multiplied by a factor of approximately 0.73, compared to districts without a PAN governor, assuming other factors are constant

(c) Provide the estimated mean number of visits from the winning PAN presidential candi- date 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
# Calculating mean number of visits from winning PAN candidate
# Given the coefficients:
# Intercept: -3.81
# competitive.district: -0.08 (for a competitive district, competitive.district=1)
# marginality.06: -2.08 (for an average poverty level, marginality.06 = 0)
# PAN.governor.06: -0.31 (for a district with a PAN governor, PAN.governor.06 = 1)

mean_visits <- exp(-3.81 + (-0.08 * 1) + (-2.08 * 0) + (-0.31 * 1))
mean_visits
```

```
## [1] 0.01499558
```

The estimated number of visits from the winning PAN presidential candidate is 0.015, given this hypothetical district is competitive, has an average poverty level and a PAN governor.