

# Gov 2001: Problem Set 6

## Instructions

You should submit your answers and R code to the problems below using the Quizzes section on Canvas.

## Problem 1

For this problem you will continue to use Model 1 from Fearon and Laitin (2003) that you coded up for Problem 1 in Problem Set 5. Make sure to fix any mistakes you made in Problem 1 in Problem Set 5 before continuing this problem.

Just as a reminder, for this problem you replicated and explored the results of the article “Ethnicity, Insurgency, and Civil War,” by James Fearon and David Laitin. This article uses a number of institutional, geographic, and social variables which have been hypothesized to predict the onset of civil conflict.

The complete dataset is available here in the .dta format. You will use the following variables in your analysis:

- onset: civil war onset that year
- warl: a ‘distinct’ civil war ongoing in previous year
- gdpenl: lagged per capita income, thousands of US dollars
- lpopl1: lagged natural logarithm of population
- lmtnest: log
- ncontig: country is non-contiguous geographically
- Oil: country is an oil exporter
- nwstate: country achieved independence in past two years
- instab: significant change in polity score in last 3 years, lagged
- polity2l: lagged level of democratization
- ethfrac: a measure of ethnic fragmentation
- relfrac: a measure of religious fragmentation

Once you have loaded the data, you will need to subset it to include only the above variables. Then delete all rows containing missing data using `na.omit()`. Finally, fix the value of onset that was incorrectly coded as a 4. You may assume for the time being that it was supposed to be a 1.

Sections 1.A to 1.C are repeated from the last problem set, but we included them here so that you can make sure to fix any mistakes you made in the previous problem set.

## 1.A

What are the stochastic and systematic components for the logit model?

## 1.B

Derive the log likelihood.

## 1.C

Replicate model 1 from Table 1, using your log likelihood from 1.B and `optim()` with *method* = “*BFGS*”. Be careful in selecting the starting values in your optimization (a vector of zeroes should be adequate). Report the coefficient and standard errors for the intercept, warl, gdpenl, polity21, and ethfrac.

## 1.D

To evaluate the role of various explanatory variables in increasing the risk of civil war, we will calculate some expected values and first differences. First, let's establish typical covariate values for states which are at risk of civil war. Using the `apply()` function, calculate the median value of all of the explanatory variables in states which experienced a civil war, using data from the year(s) of civil war onset only. Report the median values for warl, gdpenl, polity21, and ethfrac.

## 1.E

Use simulation to calculate the expected difference in the probability of civil war onset between an oil producer and a non-oil producer. This is also known as a first difference. All other variables should be set to the values you found in part 1.D. Provide a 95% confidence interval based on the appropriate quantiles of the simulations (you may find the `quantile()` or `sort()` functions useful).

## 1.F

Use simulation to calculate the expected probability of civil war onset for recently independent countries (according to the definition given by the authors) and non-recently independent countries, as lagged GDP per capita increases from \$250 to \$8000 (note that the dataset has GDP in thousands, so this is .25 to 8 in the dataset). All other covariates should be set at the levels calculated in 1.D. Plot these results as two lines on a clearly labeled graph with the probability of onset on the y-axis and GDP per capita on the x-axis.

## Problem 2

For the remainder of this problem set, we will be using a subset of data from the replication dataset for the *Logic of Political Survival* by Bruce Bueno de Mesquita, Alastair Smith, Randolph Siverson, and James Morrow. We will not be replicating their regressions, but we will be predicting the number of revolutions in a country-year observation with the following variables:

- *Revolutions*: revolutions
- *GenStrikes*: general strikes
- *GovCrises*: government crises
- *Riots*: riots
- *polity2*: polity score

The subset of their data is available [here](#).

First we will use the classic linear regression model to predict revolutions with these four variables and an intercept.

## 2.A

What are the stochastic and systematic components for the linear regression model? (Hint: you've done this before.)

## 2.B

Write out the log-likelihood function for the linear regression model.

## 2.C

Optimize the model to find the maximum likelihood estimates for the  $\beta$ s and  $\sigma$ . Report the  $\beta$  coefficients and standard errors.

## 2.D

Generate a histogram with 10,000 predicted values for a democracy (polity2=10), with all other variables held at their mean. Set the seed to 8766.

## 2.E

Based on your simulations, calculate the probability that such a state has one or more revolutions in a given year.

# Problem 3

Now we will use a Poisson model to estimate the same regression as in Problem 2.

## 3.A

What are the stochastic and systematic components for a Poisson regression model?

## 3.B

Write out the log-likelihood equation.

## 3.C

Optimize the model to find the MLE for the  $\beta$ s. Report the coefficients and standard errors.

## 3.D

Generate a histogram of 10,000 predicted values for a democracy (polity2=10), with all other variables held at their mean. Set the seed to 8766.

## 3.E

Based on your simulations, calculate the probability that such a state has one or more revolutions in a given year.

## Problem 4

The problem will use the dataset posted online called `robust.RData`.

**4.A)** Use `lm()` to run a regression of  $y$  on  $X_1$  and  $X_2$ . Report the requested coefficients and standard errors.

**4.B)** Using the output of your regression, you are now going to calculate the robust variance-covariance matrix by hand. To do this you can either:

1. Use `estfun()` to calculate the meat of the sandwich
2. Use `vcov()` to calculate the bread of the sandwich
3. Calculate your robust variance-covariance matrix
4. Check your robust standard errors with the `hccm` function using option `hc0`.

OR (you get extra points for doing it this way...)

1. Write a function that calculates the log-likelihood of each individual data point within your data. Your function should take `beta`, `X`, `y`, and `sigma`.
2. Use this function and `numericGradient()` to calculate the gradient of the log-likelihood evaluated at each point in your dataset and at the MLE for `beta` and `sigma`.
3. Use the score vector and the variance-covariance matrix from your `lm` output (you can use the function `vcov`) to calculate the robust variance-covariance matrix and robust standard errors.
4. Check your robust standard errors with the `hccm` function using option `hc0`.

Provide the code you used to calculate the robust standard errors.

**4.C)** Report the calculated robust standard errors.

**4.D)** Are your robust and regular standard errors different from one another?

**4.E)** Use residual plots to diagnose the problem with the model. Create a couple residual plots you think are interesting.

**4.F)** Alter the model to fix the misspecification. What new specification did you decide to use and why?

**4.G)** Calculate the robust and regular standard errors for your newly specified model. Do they match now? Why or why not?

## **R Code**

Please submit all your code for this assignment as a .R file. Your code should be clean, commented, and executable without error.