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 int he previous problem set.

1.A

What are the stochastic and systematic components for the logit model?
The stochastic component is $Y_i \sim Bernoulli(\pi_i)$

The systematic component is $\pi_i = \frac{1}{(1+e^{-X_i\beta})}$

1.B

Derive the log likelihood.
Derivation of the log-likelihood:

$$\begin{aligned}
L(\pi_i|y) &\propto \prod_i^N (\pi_i)^{y_i}(1-\pi_i)^{(1-y_i)} \\
lnL(\pi_i|y) &\propto \sum_i^N y_i ln(\pi_i) + (1-y_i)ln(1-\pi_i) \\
lnL(\beta|y) &\propto \sum_i^N -y_i ln(1+e^{-X_i\beta}) + (1-y_i)ln(1-\frac{1}{1+e^{-X_i\beta}}) \\
&\propto \sum_i^N -y_i ln(1+e^{-X_i\beta}) + (1-y_i)ln(\frac{e^{-X_i\beta}}{1+e^{-X_i\beta}}) \\
&\propto \sum_i^N (1-y_i)ln(e^{-X_i\beta}) - (1-y_i)ln(1+e^{-X_i\beta}) - y_i ln(1+e^{-X_i\beta}) \\
&\propto \sum_i^N (1-y_i)ln(e^{-X_i\beta}) - ln(1+e^{-X_i\beta}) + y_i ln(1+e^{-X_i\beta}) - y_i ln(1+e^{-X_i\beta}) \\
&\propto \sum_i^N (1-y_i)ln(e^{-X_i\beta}) - ln(1+e^{-X_i\beta}) \\
&\propto \sum_i^N ln(e^{-X_i\beta(1-y_i)}) - ln(1+e^{-X_i\beta})
\end{aligned}$$

Because $y_i \in \{0, 1\}$ we can show that this is equivalent to:

$$= -\sum_{i}^{N} ln(1 + e^{(1-2y_i)X_i\beta})$$

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.

	Coefficient	Std. Error
Intercept	-6.73	0.74
warl	-0.95	0.31
gdpenl	-0.34	0.07
lpopl1	0.26	0.07
lmtnest	0.22	0.08
ncontig	0.44	0.27
Oil	0.86	0.28
nwstate	1.71	0.34
instab	0.62	0.24
polity2l	0.02	0.02
ethfrac	0.17	0.37
relfrac	0.29	0.51

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.

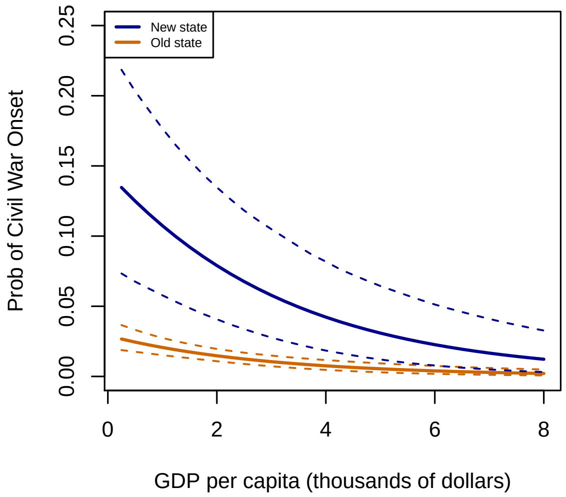
	Median
	1.00
warl	0.00
gdpenl	1.17
lpopl1	9.45
lmtnest	2.69
ncontig	0.00
Oil	0.00
nwstate	0.00
instab	0.00
polity2l	-3.00
ethfrac	0.54
relfrac	0.37

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).
The simulated 95% confidence interval for the first difference is [0.007, 0.054].

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 on the assignment page.

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.)

Stochastic component:

$$Y_i \sim N(\mu_i, \sigma^2) = f_N(y_i|\mu_i, \sigma^2)$$

Systematic component:

$$\mu_i = X_i\beta$$

2.B

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

$$\begin{aligned}
lnL(\beta,\sigma^2|y) &= ln\left[\prod \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-\mu_{it})^2}{2\sigma^2}}\right] \\
&= \sum ln\left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-X_{it}\beta)^2}{2\sigma^2}}\right] \\
&= \sum ln\left[\frac{1}{\sqrt{2\pi\sigma^2}}\right] + ln\left[e^{-\frac{(y_i-X_{it}\beta)^2}{2\sigma^2}}\right] \\
&= \sum -\frac{1}{2}ln(2\pi\sigma^2) - \frac{(y_i - X_{it}\beta)^2}{2\sigma^2} \\
&= \sum -\frac{1}{2}[ln(2\pi\sigma^2) + \frac{(y_i - X_i\beta)^2}{\sigma^2}] \\
&= \sum -\frac{1}{2}[ln(\sigma^2) + ln(2\pi) + \frac{(y_i - X_i\beta)^2}{\sigma^2}] \\
&= \sum -\frac{1}{2}[ln\sigma^2 + \frac{(y_i - X_i\beta)^2}{\sigma^2}] - \frac{1}{2}ln(2\pi) \\
&\propto \sum -\frac{1}{2}[ln\sigma^2 + \frac{(y_i - X_i\beta)^2}{\sigma^2}]
\end{aligned}$$

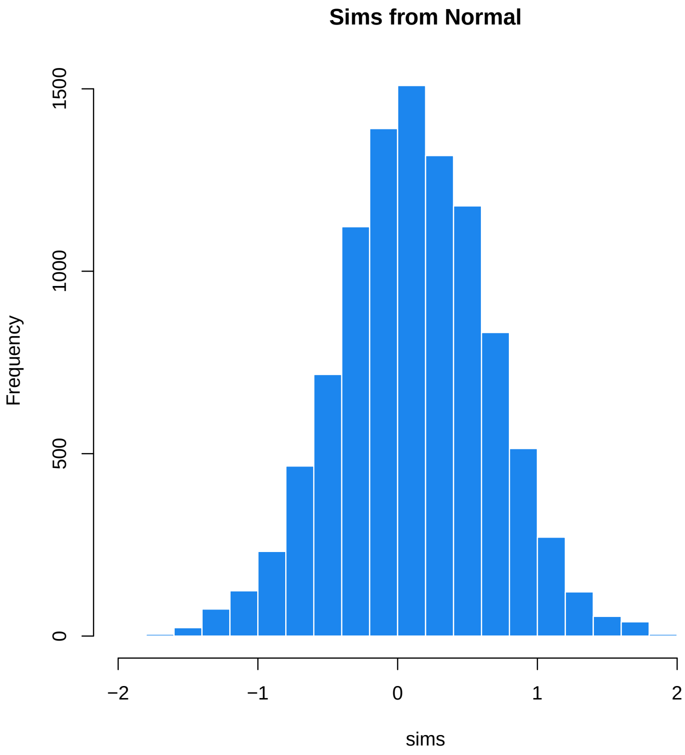
2.C

Optimize the model to find the maximum likelihood estimates for the βs and σ. Report the β coefficients and standard errors.

	Coefficient	SE
Intercept	0.15	0.01
GenStrikes	0.04	0.01
GovCrises	0.19	0.01
Riots	0.01	0.00
Polity	-0.01	0.00
sigma^2	-1.23	0.02

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.


Sims from Normal
Frequency
0
500
1000
1500
−2
−1
0
1
2
sims

2.E

Based on your simulations, calculate the probability that such a state has one or more revolutions in a given year.
The probability is simulated to be 0.0494.

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?
Stochastic component: $Y_i \sim Poisson(\lambda_i)$
Systematic component: $\lambda_i = e^{X_i \beta}$

3.B

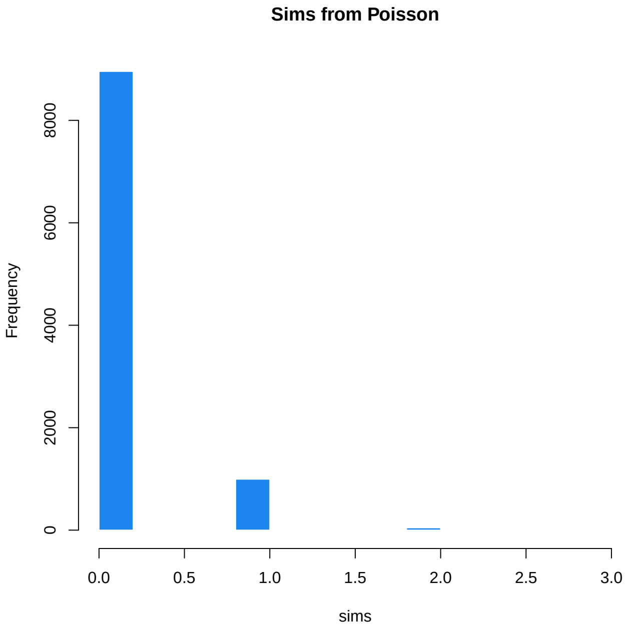
Write out the log-likelihood equation.

$$\begin{aligned}
L(\lambda_i|y_i) &= \prod_i^N \frac{\lambda^{y_i} e^{-\lambda_i}}{y_i!} \\
L(\beta_i|y_i, X_i) &\propto \prod_i^N e^{X_i\beta y_i} e^{-e^{X_i\beta}} \\
\ell(\beta_i|y_i, X_i) &\propto \sum_i^N log\big(e^{X_i\beta y_i} e^{-e^{X_i\beta}}\big) \\
&\propto \sum log\big(e^{X_i\beta y_i}\big) + log\big(e^{-e^{X_i\beta}}\big) \\
&\propto \sum X_i\beta y_i - e^{X_i\beta}
\end{aligned}$$

3.C

Optimize the model to find the MLE for the βs. Report the coefficients and standard errors.

	Coefficient	SE
Intercept	-1.83	0.03
GenStrikes	0.09	0.03
GovCrises	0.46	0.02
Riots	0.04	0.01
Polity	-0.05	0.00



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.
See histogram pictured above.

3.E

Based on your simulations, calculate the probability that such a state has one or more revolutions in a given year.
The simulated probability is 0.1044.

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.

Table 1:

X1	6.652^{***}
	(0.189)
X2	-9.350^{***}
	(0.204)
Constant	15.910^{***}
	(1.175)
Note:	*p<0.1; **p<0.05; ***p<0.01

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.

```
#First way

est.fun <- estfun(lm.1)
robust2 <- sandwich(lm.1, meat=crossprod(est.fun)/500)
robust1 <- hccm(lm.1, type="hc0")
sqrt(diag(robust1))

#Second way

ll.normal <- function(beta,y,X, sigma){
  sigma2 <- sigma
  -1/2 * (log(sigma2) + (y -(X%*%beta))^2/sigma2)
}

X <- cbind(1, X1, X2)

sigma <- sum(lm.1$residuals^2) / (nrow(model.matrix(lm.1)) -
                                    ncol(model.matrix(lm.1)))
out <- apply(cbind(y,X), 1,
             function(x) numericGradient(ll.normal,
                                         lm.1$coefficients,
                                         y=x[1],
                                         X=x[2:length(x)],
                                         sigma=sigma))

meat <- out%*%t(out)
bread <- vcov(lm.1)%*%meat%*%vcov(lm.1)
sqrt(diag(bread))
```

4.C) Report the calculated robust standard errors.

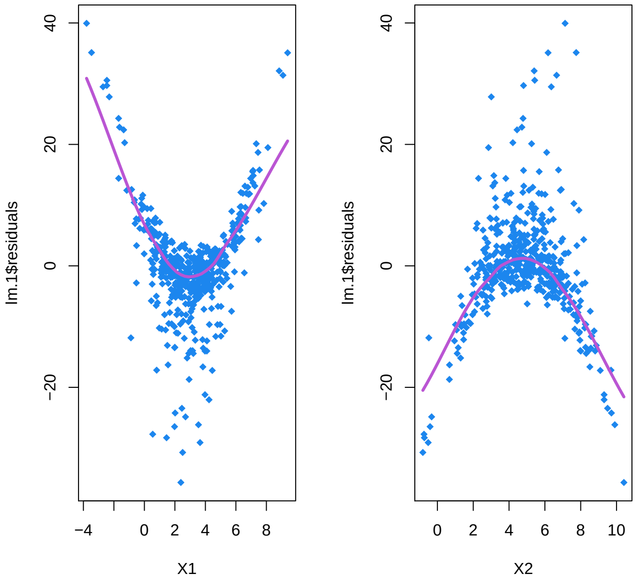
The robust standard errors are: Intercept: 1.8779; X1: 0.3444; X2:0.3416.

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

There are large differences between the robust and regular standard errors - the robust standard errors are almost twice the size of the regular standard errors.

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

Here's some nonlinearity for you:



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

Included squared terms for X1 and X2.

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

The standard errors coincide now. By included squared terms the residuals

R Code

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

	Regular	Robust
(Intercept)	0.40	0.38
X1	0.09	0.08
X2	0.16	0.15
I(X1^2)	0.01	0.01
I(X2^2)	0.02	0.02