assignment-4

Sidra Effendi

03/04/2021

```
sd(civil_war$war)
```

```
## [1] 0.3449813
```

```
mean(civil_war$war)
```

```
## [1] 0.1380514
```

Q.1

(a)

```
civil_war$ethfrac_gdp_per_capita=civil_war$ethfrac*civil_war$gdp_per_capita
model1 <- war ~ ethfrac + gdp_per_capita + ethfrac_gdp_per_capita
m1 <- lm(model1, data=civil_war) # Adjust standard errors
cov1 <- vcovHC(m1, type = "HC1")
r_se1 <- sqrt(diag(cov1)) # Robust SEs

summary(lm_robust(model1, data=civil_war, se_type = "stata"))
```

```
##
## Call:
## lm_robust(formula = model1, data = civil_war, se_type = "stata")
##
## Standard error type:  HC1
##
## Coefficients:
##                          Estimate Std. Error t value  Pr(>|t|)  CI Lower
## (Intercept)               0.10242   0.009570  10.702 1.677e-26  0.083656
## ethfrac                   0.19548   0.021098   9.265 2.614e-20  0.154123
## gdp_per_capita           -0.00646   0.001032  -6.257 4.182e-10 -0.008484
## ethfrac_gdp_per_capita -0.01507   0.002769  -5.443 5.444e-08 -0.020499
##                           CI Upper   DF
## (Intercept)               0.121175 6298
## ethfrac                   0.236842 6298
## gdp_per_capita           -0.004436 6298
## ethfrac_gdp_per_capita -0.009643 6298
```

```
##
## Multiple R-squared:  0.04619 ,   Adjusted R-squared:  0.04573
## F-statistic: 140.4 on 3 and 6298 DF,  p-value: < 2.2e-16
```

The coefficient on the interaction term is -0.01507 and it is statistically significant since the p-value is less than 0.05 and |t-value|=5.443 is greater than 1.96. Compared to the standard deviation on the output column war which is 0.345, the magnitude of interaction term coefficient is very small.

To interpret the coefficient on interaction term we can consider there is unit change in the GDP, that is,

delta (war)/delta(GDP) = beta_{gdp_per_capita} + beta_{ethfrac_gdp_per_capita} * ethfrac

```
= -0.00646 -0.01507 * ethfrac
```

We can plug in value for ethfrac= 0.5960, which I picked from the given dataset.

delta (war)/delta(GDP) = -0.00646 -0.01507 * 0.5960 = -0.0154

On average, we predict there is a 1.54 percentage points decrease in the probability that there is a interstate conflict in a country associated with a unit change in the GDP per capita of a country when probability that two randomly drawn individuals in a country will be from different ethnolinguistic groups (measure of ethnic fractionalization) is 0.5960.

To interpret the coefficient on interaction term we can also consider there is unit change in the measure of ethnic fractionalization, that is,

delta(war)/delta(ethfrac) = beta_{etcfrac} + beta_{ethfrac_gdp_per_capita} * gdp_per_capita

```
= 0.19548 -0.01507 * gdp_per_capita
```

We can plug in value for gdp_per_capita= 21.0992527, which I picked from the given dataset.

delta (war)/delta(ethfrac) = 0.19548 - 0.01507 * 21.0992527 = -0.1225

On average, we predict there is a 12.25 percentage points decrease in the probability that there is a interstate conflict in a country associated with a unit change in the probability that two randomly drawn individuals in a country will be from different ethnolinguistic groups when the GDP per capita of a country is 0.6312.

Q.2

(a)

```
civil_war$ethfrac_ncontig = civil_war$ethfrac*civil_war$ncontig
model2 <- war ~ ethfrac + ncontig + ethfrac_ncontig
m2 <- lm(model2, data=civil_war) # Adjust standard errors
cov2   <- vcovHC(m2, type = "HC1")
r_se2 <- sqrt(diag(cov2)) # Robust SEs

summary(lm_robust(model2, data=civil_war, se_type = "stata"))
```

```
##
## Call:
## lm_robust(formula = model2, data = civil_war, se_type = "stata")
```

```
##
## Standard error type:  HC1
##
## Coefficients:
##                 Estimate Std. Error t value  Pr(>|t|) CI Lower  CI Upper   DF
## (Intercept)      0.06405   0.006482   9.881 7.376e-23  0.05134 0.0767549 6298
## ethfrac          0.13258   0.015506   8.550 1.524e-17  0.10218 0.1629764 6298
## ncontig         -0.03372   0.017699  -1.905 5.684e-02 -0.06841 0.0009812 6298
## ethfrac_ncontig  0.46126   0.046231   9.977 2.849e-23  0.37063 0.5518892 6298
##
## Multiple R-squared:  0.07029 ,   Adjusted R-squared:  0.06985
## F-statistic: 104.7 on 3 and 6298 DF,  p-value: < 2.2e-16
```

The coefficient on the interaction term is 0.46126 and it is statistically significant since the p-value is quite less than 0.05 and |t-value| = 9.977 is much greater than 1.96. Compared to the standard deviation on the output column war which is 0.345, the magnitude of interaction term coefficient is large.

Our regression equation is,

war^ = 0.06405 - 0.13258* ethfrac - 0.03372 * ncontig + 0.46126 * ethfrac * ncontig

Regression equation for non-contiguous countries,

war1^ = 0.06405 - 0.13258* ethfrac - 0.03372 * (1) + 0.46126 * ethfrac * (1) = 0.03033 + 0.32868* ethfrac

Regression equation for contiguous countries,

war2^ = 0.06405 - 0.13258* ethfrac - 0.03372 * (0) + 0.46126 * ethfrac * (0) = 0.06405 - 0.13258* ethfrac

For ethfrac = 0.5960 (picked from dataset),

war1^ = 0.226

war2^ = -0.015

For ethfrac = 0.5960, we observer the estimated average probability of interstate conflict in a county to increase by 22.6 percentage points for a non-contiguous country (country with territory holding at least 10,000 people and separated from the land area containing the capital city either by land or by 100 km of water), where as for a contiguous country the estimated average probability of interstate conflict will decrease by 1.5 percentage points.

Q.3

(a)

```
model3 <- war ~ log_lagged_pop + log_lagged_gdp_per_capita + ncontig + ethfrac + new_state + dem_high +
m3 <- lm(model3, data=civil_war) # Adjust standard errors
cov3 <- vcovHC(m3, type = "HC1")
r_se3 <- sqrt(diag(cov3)) # Robust SEs

summary(lm_robust(model3, data=civil_war, se_type = "stata"))
```

```
##
## Call:
## lm_robust(formula = model3, data = civil_war, se_type = "stata")
```

```
##
## Standard error type:  HC1
##
## Coefficients:
##                             Estimate Std. Error t value   Pr(>|t|)   CI Lower
## (Intercept)                  0.03497   0.038931  0.8983  3.691e-01 -0.0413476
## log_lagged_pop               0.03203   0.002762 11.5963  8.832e-31  0.0266183
## log_lagged_gdp_per_capita -0.04131   0.004213 -9.8041  1.566e-22 -0.0495679
## ncontig                      0.07670   0.010976  6.9873  3.090e-12  0.0551782
## ethfrac                      0.06743   0.012439  5.4207  6.156e-08  0.0430429
## new_state                    0.04220   0.022747  1.8550  6.364e-02 -0.0023959
## dem_high                     0.01711   0.008222  2.0814  3.744e-02  0.0009954
## instab                      -0.01060   0.010717 -0.9895  3.225e-01 -0.0316147
## prior                        0.64023   0.016055 39.8767 3.089e-310  0.6087576
##                             CI Upper   DF
## (Intercept)                  0.11129 6293
## log_lagged_pop               0.03745 6293
## log_lagged_gdp_per_capita -0.03305 6293
## ncontig                      0.09821 6293
## ethfrac                      0.09181 6293
## new_state                    0.08679 6293
## dem_high                     0.03323 6293
## instab                       0.01040 6293
## prior                        0.67171 6293
##
## Multiple R-squared:  0.4995 ,    Adjusted R-squared:  0.4988
## F-statistic: 343.7 on 8 and 6293 DF,  p-value: < 2.2e-16
```

The order is similar to one in the above output.

1. A 1% increase in the population in thousands in the country in the lagged year is on average associated with a 0.032 (0.03203 * 0.01) percentage points or 23.2 [(0.03203/0.1380514) *100] percent increase in the probability that there will be interstate conflict in the country, holding lagged_gdp_per_capita, ncontig, ethfrac, new_state,dem_high,instab and prior as constant. The coefficient is statistically significant at 5% confidence as t-value=11.5963 is greater than 1.96 and p-value is less than 0.05. Also, statistically significant at 10% and 1% levels.

2. A 1% increase in the lagged gross domestic product (GDP) per capita of a country is on average associated with a 0.041 (0.04131 * 0.01) percentage points or 29.9 [(0.04131/0.1380514) *100] percent decrease in the probability that there will be interstate conflict in the country, holding other variables constant. The coefficient is statistically significant at 5% confidence as t-value=|-9.8041| is greater than 1.96 and p-value is less than 0.05. Also, statistically significant at 10% and 1% levels.

3. On average the probability that there will be interstate conflict in a non-contiguous country is predicted to be 7.67 percentage points or 55.5 [(0.0767/0.1380514) *100] percent more than if the country was contiguous, holding other variables constant. The coefficient is statistically significant at 5% confidence as t-value=|6.9873| is greater than 1.96 and p-value is less than 0.05. Also, statistically significant at 10% and 1% levels.

4. On average, with one unit increase in the probability that two randomly drawn individuals in a country will be from different ethnolinguistic groups, the probability that there will be interstate conflict in the country is predicted to increase by 6.74 percentage points or 48.8 [(0.06743/0.1380514) *100] percent, holding other variables constant. The coefficient is statistically significant at 5% confidence as t-value=|5.4207| is greater than 1.96 and p-value is less than 0.05. Also, statistically significant at 10% and 1% levels.

5. On average the probability that there will be interstate conflict in a country in the first year of it's existence is predicted to be 4.22 percentage point or 30.57 [(0.04220/0.1380514)*100] percent more than if the country was not in the first year of it's existence, holding other variables constant. The coefficient is statistically not significant at 5% (and 1%) confidence level as p-value is greater than 0.05 and t-value is less than 1.96, but, statistically significant at 10% as p-value < 0.1.

6. On average the probability that there will be interstate conflict in a country where the level of democracy is high or the level of democracy in a country in the lagged year is greater than 5 is predicted to be 1.71 percentage point or 12.4 [(0.01711/0.1380514) *100] percent more than if the country does not have a high level of democracy, holding other variables constant. The coefficient is statistically significant at 5% confidence level as p-value is less than 0.05 and t-value= 2.0814 is greater than 1.96. Also, statistically significant at 10% but not at 1%.

7. On average the probability that there will be interstate conflict in a country which is unstable or the level of democracy in the country changed by at least 2 in the last 3 years in the lagged year, is predicted to be 1.06 percentage point or 7.65 [(0.01060/0.1380514) *100] percent less than if the country was not unstable, holding other variables constant. The coefficient is not statistically significant at 5% confidence level as p-value is greater than 0.05 and t-value is less than 1.96. Also, not statistically significant at 10% and 1% levels.

8. On average the probability that there will be interstate conflict in a country in the given year if there was an ongoing civil conflict in the country in the prior year, is predicted to be 64.023 percentage point or 46.4 [(0.64023/0.1380514) *100] percent more than if the country had seen no ongoing civil conflict in the previous year, holding other variables constant. The coefficient is statistically significant at 5% confidence level as p-value is less than 0.05 and t-value=39.8767 is greater than 1.96. Also, statistically significant at 10% and 1% levels.

Q.4

```
yhat <- as_tibble(predict(m3))
yhat %>% mutate(out=if_else(value>1 | value < 0 ,1,0)) %>%
summarise(perc_out=mean(out))
```

```
## # A tibble: 1 x 1
##   perc_out
##      <dbl>
## 1    0.238
```

We see that using the LPM model we have 23.8% output values which are either less than 0 or greater than 1. We want the probability of interstate conflict in a country in a given year and any value outside the bounds of (0,1) is meaningless and might lead to confusion.

Q.5

```
m4 <- glm(model3, family = "binomial", data=civil_war)
# Adjust standard errors
cov4  <- vcovHC(m4, type = "HC1")
r_se4 <- sqrt(diag(cov4)) # Robust SEs
```

(a)

```
yhat <-as_tibble(predict(m4, newdata = civil_war, type = "response"))
yhat %>% mutate(out=if_else(value>1 | value < 0 ,1,0)) %>%
summarise(perc_out=mean(out))
```

```
## # A tibble: 1 x 1
##   perc_out
##      <dbl>
## 1        0
```

Here we see that we get no values less than 0 or greater than 1, for probability of interstate conflict in a country in a given year. This is because we are using a logit model which bounds the predicted values between 0 and 1.

(b)

```
#summary(lm_robust(model3, data=civil_war, se_type = "stata"))
summary(m4)
```

```
##
## Call:
## glm(formula = model3, family = "binomial", data = civil_war)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.8464  -0.3147  -0.2027  -0.1269   3.3022
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -3.19941    0.61635  -5.191 2.09e-07 ***
## log_lagged_pop             0.50548    0.04067  12.429  < 2e-16 ***
## log_lagged_gdp_per_capita -0.71558    0.06460 -11.077  < 2e-16 ***
## ncontig                    0.98701    0.14739   6.697 2.13e-11 ***
## ethfrac                    0.82190    0.19579   4.198 2.69e-05 ***
## new_state                  0.66372    0.31471   2.109   0.0349 *
## dem_high                   0.27179    0.13604   1.998   0.0457 *
## instab                    -0.06037    0.13834  -0.436   0.6625
## prior                      4.10144    0.12029  34.097  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5059.4  on 6301  degrees of freedom
## Residual deviance: 2533.5  on 6293  degrees of freedom
## AIC: 2551.5
##
## Number of Fisher Scoring iterations: 6
```

Comparing LPM to the logit model, we see that as per the logit model the variable dem_high is statistically significant only at 10% level whereas in LPM dem_high is also statistically significant at 5% level. In terms of sign only the sign of the intercept has changed from +ve to -ve and it wasn't significant in the LPM model not even at 10% level but is significant in logit model at 1%, 5% and 10% level. Rest the significance values and signs are similar for other variables in both the models.

Q.6

Our regression equation from question3 for LPM model is

warˆ = 0.03497+ 0.03203* log_lagged_pop-0.04131* log_lagged_gdp_per_capita+0.07670* ncontig+0.06743* ethfrac+ 0.04220* new_state + 0.01711* dem_high -0.01060 * instab+ 0.64023 *prior

For the given values,

war1ˆ = 0.03497+ 0.03203* 8 -0.04131* 5 +0.07670* 1 +0.06743* 0.5+ 0.04220* 1 + 0.01711* 0 -0.01060 * 0+ 0.64023 *0

```
= 0.0443825 -0.04131* 5  = 0.237275
```

After lagged GDP per capita increased by 10%,

war2ˆ = 0.03497+ 0.03203* 8 -0.04131* 5.1 +0.07670* 1 +0.06743* 0.5+ 0.04220* 1 + 0.01711* 0 -0.01060 * 0+ 0.64023 *0

```
= 0.0443825 -0.04131* 5.1 = 0.233144
```

The change in probability is,

war2^{-war1} = 0.233144 - 0.237275 = -0.004131

If the lagged gdp per capita in the country increased by 10% there is a 0.41 percentage point decrease in the probability that there will be a interstate conflict in a country, holding other at variables constant value.

Q.7

(a)

Our regression equation for logit model is

F = -3.19941+ 0.50548* log_lagged_pop-0.71558* log_lagged_gdp_per_capita+0.98701* ncontig+0.82190* ethfrac+ 0.66372* new_state + 0.27179* dem_high -0.06037 * instab+ 4.10144 *prior

For the given values,

F1 = -3.19941 + 0.50548* 8 - 0.71558* 5 + 0.98701* 1 + 0.82190* 0.5+ 0.66372* 1 + 0.27179* 0 -0.06037 * 0 + 4.10144 *0

```
= - 0.67179
```

After agged GDP per capita increased by 10%,

F2 = -3.19941 + 0.50548* 8 - 0.71558* 5.1 + 0.98701* 1 + 0.82190* 0.5+ 0.66372* 1 + 0.27179* 0 -0.06037 * 0 + 4.10144 *0

```
= -0.74338
```

(b)

The change in probability is,

Pr(war=1) = 1/[1+e^(F1)] = 0.3381

Pr(war=1) = 1/[1+e^(F2)] = 0.3223

0.3381 - 0.3223 = 0.0158

If the lagged gdp per capita in a country increased by 10% there is a 1.58 percentage point decrease in the probability that there will be a interstate conflict in a country, holding other at variables constant value.

(c)

```
at <- data.frame(log_lagged_pop=8,

                 log_lagged_gdp_per_capita=c(5,5.1),

                 ncontig=1,

                 ethfrac=0.5,

                 new_state=1,

                 dem_high=0,

                 instab=0,

                 prior=0)


margins1 <- margins(m4, at = at,type = "response")

margins1 %>% group_by(log_lagged_pop, log_lagged_gdp_per_capita, ncontig, ethfrac, new_state, dem_high,

                 summarize(margin=mean(fitted)) %>% ungroup() %>%

                 dplyr::select(log_lagged_gdp_per_capita,margin)
```

```
## `summarise()` has grouped output by 'log_lagged_pop', 'log_lagged_gdp_per_capita', 'ncontig', 'ethfra
```

```
## # A tibble: 2 x 2
##   log_lagged_gdp_per_capita margin
##                       <dbl>  <dbl>
## 1                         5   0.338
## 2                         5.1 0.322
```

If the lagged gdp per capita in a country increased by 10% there is a 1.58 percentage point decrease in the probability that there will be a interstate conflict in a country, holding other at variables constant value.

Q.8

For the same set set of values and same % increase in the value of log gdp per capita, LPM reported a decrease in the probability of interstate conflict in a country as 0.41 percentage points while as logit reported a decrease of 1.58 percentage points. We know for sure that both of these model fit the data in a different way and that is why we get such different results. LPM results in a linear line fitted on the data while as logit fits a S-shaped curve on the data.

```
stargazer(m1,m2,m3,m4, se=list(r_se1,r_se2,r_se3,r_se4), type = "text", header=TRUE, omit.stat = c("f",
```

```
##
## Table 1: Results of Regression of annual earnings on years of education and other control variables
## ================================================================
##                                          war
##                                        OLS               logistic
##                              (1)        (2)       (3)       (4)
## ----------------------------------------------------------------
## log_lagged_pop                                     0.03***   0.51***
##                                                    (0.003)   (0.04)
##
## log_lagged_gdp_per_capita                         -0.04***  -0.72***
##                                                    (0.004)   (0.07)
##
## ethfrac                     0.20***    0.13***    0.07***   0.82***
##                              (0.02)    (0.02)     (0.01)    (0.20)
##
## gdp_per_capita              -0.01***
##                             (0.001)
##
## ethfrac_gdp_per_capita      -0.02***
##                             (0.003)
##
## new_state                                           0.04*     0.66*
##                                                    (0.02)    (0.34)
##
## dem_high                                           0.02**     0.27*
##                                                    (0.01)    (0.14)
##
## instab                                              -0.01     -0.06
##                                                    (0.01)    (0.15)
##
## prior                                              0.64***   4.10***
##                                                    (0.02)    (0.13)
##
## ncontig                                 -0.03*     0.08***   0.99***
##                                         (0.02)     (0.01)    (0.17)
##
## ethfrac_ncontig                         0.46***
##                                         (0.05)
##
## Constant                    0.10***    0.06***     0.03     -3.20***
##                              (0.01)    (0.01)     (0.04)    (0.69)
##
```

```
## ------------------------------------------------------------------
## Observations                6,302      6,302     6,302     6,302
## R2                           0.05       0.07      0.50
## Adjusted R2                  0.05       0.07      0.50
## Log Likelihood                                            -1,266.77
## Akaike Inf. Crit.                                         2,551.54
## ==================================================================
## Note:                                  *p<0.1; **p<0.05; ***p<0.01
##                            Robust standard errors in parantheses
```