Quantitative and Formal Methods in IR
Class 4: Interactions and Logit

Dave Armstrong
University of Oxford
Department of Politics and International Relations
Center for Research Methods in the Social Sciences
Nuffield College

March 8, 2009

For today's class, we will be using two datasets. The first is the one we used last week, for class 3 and the other one is the one for today: class4.dta. Both can be downloaded from my website, `http://www.quantoid.net/IRStats.php`.

1 Interactions and Conditional Hypotheses

Remember, that we have talked about adding dummy variables for civilizations groups that change the intercept, but the effect of log(GDP/capita) is not changed - it has the same coefficient for each civilization. We can relax this assumption and allow the coefficients to change across the civilizations. We are then getting estimates one regression per civilization as follows:

$$
\begin{aligned}
PR_O &= a_O + b_O log(GDP/capita)_O \\
PR_A &= a_A + b_A log(GDP/capita)_A \\
PR_I &= a_I + b_I log(GDP/capita)_I \\
PR_L &= a_L + b_L log(GDP/capita)_L \\
PR_{Or} &= a_{Or} + b_{Or} log(GDP/capita)_{Or} \\
PR_W &= a_W + b_W log(GDP/capita)_W
\end{aligned}
$$

To do this, we need to make a bunch of new variables. We need to make a number of *interaction* terms where we multiply each of the civilization dummy variables by the `log_gdppc` variable. I've done this already in the dataset, but this is what the commands would look like:

```
gen afrxgdp = africa*log_gdppc
gen islxgdp = islamic*log_gdppc
gen latxgdp = latinam*log_gdppc
gen orthxgdp = orthodox*log_gdppc
gen westxgdp = western*log_gdppc
```

Now, we simply need to include all of these new variables, all of the civilization dummies and the log of GDP/capita in the model as follows:

```
. reg new_polrts african islamic latinam orthodox western log_gdppc
    afrxgdp islxgdp latxgdp orthxgdp westxgdp

      Source |       SS       df       MS              Number of obs =     176
-------------+------------------------------           F( 11,   164) =   17.98
       Model |  86.3882947    11  7.85348134           Prob > F      =  0.0000
    Residual |  71.6517605   164  .436900979           R-squared     =  0.5466
-------------+------------------------------           Adj R-squared =  0.5162
       Total |  158.040055   175   .90308603           Root MSE      =  .66098

------------------------------------------------------------------------------
  new_polrts |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     african |   2.993281   1.262939     2.37   0.019     .4995646    5.486997
     islamic |   5.377771   1.184767     4.54   0.000     3.038407    7.717134
     latinam |   1.166195   3.055931     0.38   0.703    -4.867847    7.200237
    orthodox |   2.637315   3.009947     0.88   0.382     -3.30593    8.580559
     western |   3.663915   1.941193     1.89   0.061    -.1690377    7.496868
   log_gdppc |    .545989   .1018422     5.36   0.000      .344898    .7470801
     afrxgdp |  -.3883034   .1564956    -2.48   0.014    -.6973093   -.0792974
     islxgdp |   -.747072    .139015    -5.37   0.000    -1.021562    -.472582
     latxgdp |  -.1185928   .3472236    -0.34   0.733    -.8041978    .5670121
    orthxgdp |  -.3449979   .3335723    -1.03   0.303    -1.003648    .3136523
    westxgdp |  -.3602539   .2014846    -1.79   0.076    -.7580922    .0375844
       _cons |  -4.548367   .8689226    -5.23   0.000    -6.264084   -2.832649
------------------------------------------------------------------------------
```

This is just a bit more complicated version of what we did last time. However, just think of this as another plugging in numbers problem. For the "other" civilization, all of the dummy variables will be zero, which leaves:

$$-4.55 + 0.55log(GDP/capita)$$

For the African civilization, the `africa` variable gets a 1 and all other civilization dummies get a value of zero. This leaves:

$$-4.55+2.99+0.55log(GDP/capita)-0.39log(GDP/capita) = -1.56+(0.55-0.39)log(GDP/capita)$$

We could write out all of the other equations, too:

Other	-4.55 + 0.55 log(GDP/capita)	-4.55 + 0.55 log(GDP/capita)
African	-4.55 + 2.99 + 0.55log(GDP/capita) - 0.39log(GDP/capita)	-1.56 + 0.16log(GDP/capita)
Islamic	-4.55 + 5.38 + 0.55log(GDP/capita) - 0.75log(GDP/capita)	0.83 - 0.2log(GDP/capita)
Latin Am	-4.55 + 1.17 + 0.55log(GDP/capita) - 0.12log(GDP/capita)	-3.38 + 0.43log(GDP/capita)
Orthodox	-4.55 + 2.64 + 0.55log(GDP/capita) - 0.34log(GDP/capita)	-1.91 + 0.21log(GDP/capita)
Western	-4.55 + 3.66 + 0.55log(GDP/capita) - 0.36log(GDP/capita)	-0.89 + 0.19log(GDP/capita)

So, we basically have a different equation for each civilization. These are what we refer to as conditional coefficients and conditional intercepts. This just means that the intercept and slope of `log_gdppc` is conditional on some other variable - civilizations. What if we want to know whether the conditional coefficient is significant. For instance, let's look at the conditional coefficient for the African civilization.

```
 lincom log_gdppc + afrxgdp

 ( 1)  log_gdppc + afrxgdp = 0
```

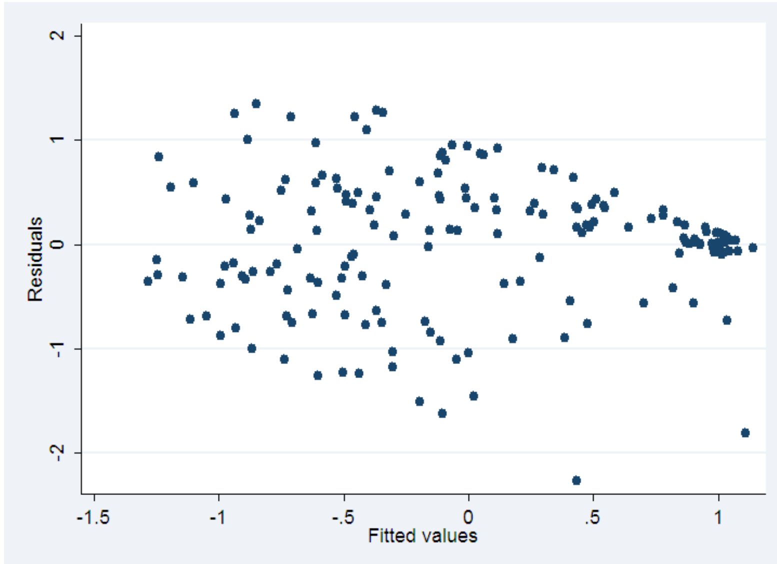
```
------------------------------------------------------------------------------
  new_polrts |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .1576857   .1188235     1.33   0.186    -.0769354    .3923068
------------------------------------------------------------------------------
```

This tells us that the conditional coefficient for `log_gdppc` for the African civilization is -.16. Since the p-value is bigger than 0.05, we cannot reject the null that this conditional coefficient is zero. That means that in Africa, political rights are not well explained by GDP/capita. We could do this for the other coefficients as well:

```
. lincom log_gdppc + islxgdp

 ( 1)  log_gdppc + islxgdp = 0

------------------------------------------------------------------------------
  new_polrts |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   -.201083   .0946221    -2.13   0.035    -.3879176   -.0142484
------------------------------------------------------------------------------

. lincom log_gdppc + latxgdp

 ( 1)  log_gdppc + latxgdp = 0

------------------------------------------------------------------------------
  new_polrts |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .4273962   .3319524     1.29   0.200    -.2280552    1.082848
------------------------------------------------------------------------------

. lincom log_gdppc + orthxgdp

 ( 1)  log_gdppc + orthxgdp = 0

------------------------------------------------------------------------------
  new_polrts |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .2009912   .3176455     0.63   0.528    -.4262109    .8281932
------------------------------------------------------------------------------

. lincom log_gdppc + westxgdp

 ( 1)  log_gdppc + westxgdp = 0

------------------------------------------------------------------------------
  new_polrts |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .1857351   .1738511     1.07   0.287    -.1575399    .5290102
------------------------------------------------------------------------------
```

The only significant conditional coefficient is the one for the Islamic world - there, as GDP/capita increases, political rights get worse. How would we do this for the intercepts?

Now, let's go ahead and look at the residuals again:



This shows that again, the residuals look as though the mean is everywhere zero, perhaps more so than in the previous plot, however the fan-shape indicating heteroskedasticity is still a problem. Let's re-estimate the model above with robust standard errors and then re-calculate the conditional coefficients and their standard errors.

```
. reg new_polrts african islamic latinam orthodox western log_gdppc afrxgdp islxgdp latxgdp orthxgdp w
> estxgdp, robust

Linear regression                                      Number of obs =     176
                                                       F( 11,   164) =   46.31
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.5466
                                                       Root MSE      =  .66098

------------------------------------------------------------------------------
             |               Robust
  new_polrts |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     african |   2.993281   1.543228     1.94   0.054    -.0538772    6.040438
     islamic |   5.377771   1.387917     3.87   0.000     2.637281    8.118261
     latinam |   1.166195   1.710824     0.68   0.496    -2.211887    4.544277
    orthodox |   2.637315   3.428734     0.77   0.443     -4.13284    9.407469
     western |   3.663915   1.332659     2.75   0.007     1.032534    6.295296
   log_gdppc |    .545989   .1296443     4.21   0.000      .290002    .8019761
     afrxgdp |  -.3883034   .1946084    -2.00   0.048    -.7725645   -.0040422
     islxgdp |   -.747072   .1598509    -4.67   0.000    -1.062703    -.431441
     latxgdp |  -.1185928   .2010911    -0.59   0.556    -.5156542    .2784685
    orthxgdp |  -.3449979   .3965785    -0.87   0.386    -1.128056    .4380601
    westxgdp |  -.3602539   .1478462    -2.44   0.016    -.6521813   -.0683265
       _cons |  -4.548367   1.122128    -4.05   0.000    -6.764047   -2.332686
------------------------------------------------------------------------------

.
end of do-file

. lincom log_gdppc + afrxgdp

 ( 1)  log_gdppc + afrxgdp = 0

------------------------------------------------------------------------------
  new_polrts |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .1576857   .1451372     1.09   0.279    -.1288928    .4442641
------------------------------------------------------------------------------

. lincom log_gdppc + islxgdp
```

```
 ( 1)  log_gdppc + islxgdp = 0

------------------------------------------------------------------------------
  new_polrts |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   -.201083   .0935129    -2.15   0.033    -.3857275   -.0164385
------------------------------------------------------------------------------

. lincom log_gdppc + latxgdp

 ( 1)  log_gdppc + latxgdp = 0

------------------------------------------------------------------------------
  new_polrts |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .4273962   .1537205     2.78   0.006     .1238698    .7309226
------------------------------------------------------------------------------

. lincom log_gdppc + orthxgdp

 ( 1)  log_gdppc + orthxgdp = 0

------------------------------------------------------------------------------
  new_polrts |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .2009912   .3747891     0.54   0.592    -.5390429    .9410252
------------------------------------------------------------------------------

. lincom log_gdppc + westxgdp

 ( 1)  log_gdppc + westxgdp = 0

------------------------------------------------------------------------------
  new_polrts |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .1857351   .0710694     2.61   0.010     .0454062    .3260641
------------------------------------------------------------------------------
```

Now that we've dealt appropriately with the standard errors, you can see that three of the conditional coefficients are statistically significant - the Islamic, Latin American and Western coefficients are all statistically significantly different from zero.

2 Logistic Regression

There are lots of important aspects of the logistic regression model that we are simply not going to have time to discuss. What we are going to discuss is how to interpret the model results. There are other ways to interpret the coefficients from this model, but the one I propose here is the most intuitive way and therefore should be easiest to understand.

Remember from regression what we want to know is how much the probability of observing a one on the dependent variable changes as a function of a change in the independent variable. As we suggested in the lecture, the effect of a one-unit change in an independent variable will have different effects on the change in probabilities because of the difference in the marginal effects.

Let's look at the following model where we will predict whether or not a country is labeled as "free" by Freedom House as a function of GDP/capita. We could write the model in math as follows:

$$\log\left(\frac{Pr(Y=1)}{1-Pr(Y=1)}\right) = a + b_1 \log(GDP/capita)$$

or as

$$Pr(Y=1) = \frac{1}{1+exp(-(a+b_1\log(GDP/capita)))}$$

We cannot interpret the coefficients as we did with linear regression. Instead, we simply have to calculate predicted probabilities. To get the probability that a country is free, we simply make the calculating in the second of the expressions above:

```
. logit free log_gdppc

Iteration 0:   log likelihood =  -120.0664
Iteration 1:   log likelihood =  -88.99596
Iteration 2:   log likelihood = -86.845056
Iteration 3:   log likelihood = -86.766896
Iteration 4:   log likelihood = -86.766742

Logistic regression                               Number of obs   =        176
                                                  LR chi2(1)      =      66.60
                                                  Prob > chi2     =     0.0000
Log likelihood = -86.766742                       Pseudo R2       =     0.2773

------------------------------------------------------------------------------
        free |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   log_gdppc |   1.384591   .2152951     6.43   0.000     .9626207    1.806562
       _cons |  -12.58186    1.94596    -6.47   0.000    -16.39587   -8.767848
------------------------------------------------------------------------------
```

Let's think about the probability of being free for a country with a GDP/capita of 1000:

$$\begin{aligned}
Pr(Free=1) &= \frac{1}{1+exp(-(-12.6+1.4*\log(GDP/capita)))} \\
&= \frac{1}{1+exp(-(-12.6+1.4*\log(1000)))} \\
&= \frac{1}{1+exp(-(-12.6+1.4*6.91))} \\
&= \frac{1}{1+exp(-(-3.01))} \\
&= \frac{1}{1+20.29} \\
&= 0.047
\end{aligned}$$

Now, let's look at the probability of a country being free with a GDP/capita of 2000:

$$
\begin{aligned}
Pr(Free = 1) &= \frac{1}{1 + exp(-(-12.6 + 1.4 * \log(GDP/capita)))} \\
&= \frac{1}{1 + exp(-(-12.6 + 1.4 * \log(2000)))} \\
&= \frac{1}{1 + exp(-(-12.6 + 1.4 * 7.6))} \\
&= \frac{1}{1 + exp(-(-1.96))} \\
&= \frac{1}{1 + 7.10} \\
&= 0.123
\end{aligned}
$$

So, increasing GDP/capita from \$1000 to \$2000 increases the probability of being free by 0.123-0.047 = 0.076. This is a pretty big gain in the probability of being free. Let's look at the differences from \$9000 to \$10000 and \$50000 to \$51000:

$$
\begin{aligned}
Pr(Free = 1|GDP/capita = 9000) &= \frac{1}{1 + exp(-(-12.6 + 1.4 * \log(9000)))} \\
&= \frac{1}{1 + exp(-(-12.6 + 1.4 * 9.10))} \\
&= 0.537 \\
Pr(Free = 1|GDP/capita = 10000) &= \frac{1}{1 + exp(-(-12.6 + 1.4 * \log(10000)))} \\
&= \frac{1}{1 + exp(-(-12.6 + 1.4 * 9.21))} \\
&= 0.573 \\
Pr(Free = 1|GDP/capita = 50000) &= \frac{1}{1 + exp(-(-12.6 + 1.4 * \log(50000)))} \\
&= \frac{1}{1 + exp(-(-12.6 + 1.4 * 10.82))} \\
&= 0.927 \\
Pr(Free = 1|GDP/capita = 51000) &= \frac{1}{1 + exp(-(-12.6 + 1.4 * \log(51000)))} \\
&= \frac{1}{1 + exp(-(-12.6 + 1.4 * 10.84))} \\
&= 0.929
\end{aligned}
$$

The effect of a \$1000 increase in GDP/capita is as follows:

Change in GDP/capita	Δ Pr(Free)
\$1000→\$2000	0.076
\$9000→\$10000	0.036
\$50000→\$51000	0.002

One-thousand dollars has a different effect depending on where you start, just as in the linear regression with the logged version of GDP/capita.

Now, let's add in the civilization dummy variables to see whether they add something to the model:

```
. logit free log_gdppc african-western

Iteration 0:   log likelihood =  -120.0664
Iteration 1:   log likelihood = -68.047485
Iteration 2:   log likelihood =  -59.82407
Iteration 3:   log likelihood = -57.959211
Iteration 4:   log likelihood = -57.763806
Iteration 5:   log likelihood = -57.759807
Iteration 6:   log likelihood = -57.759804

Logistic regression                               Number of obs   =        176
                                                  LR chi2(6)      =     124.61
                                                  Prob > chi2     =     0.0000
Log likelihood = -57.759804                       Pseudo R2       =     0.5189

------------------------------------------------------------------------------
        free |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   log_gdppc |   1.462572   .3073795     4.76   0.000     .8601192    2.065025
     african |  -.6797471   .6848564    -0.99   0.321    -2.022041    .6625468
     islamic |  -4.720045   1.188016    -3.97   0.000    -7.048513   -2.391576
     latinam |  -.7534577   .6221727    -1.21   0.226    -1.972894    .4659783
    orthodox |  -1.812512   .7875062    -2.30   0.021    -3.355995   -.2690277
     western |   2.019833   1.187144     1.70   0.089    -.3069266    4.346592
       _cons |  -12.36677   2.621613    -4.72   0.000    -17.50503   -7.228501
------------------------------------------------------------------------------
```

Now, let's imagine what the biggest possible change in GDP/capita could do for each type of civilization. To do this, we have to calculate two different predicted probabilities holding each civilization dummy at one, in turn. Let's start with the "others" (where all of the included dummy variables are all set to zero):

$$\begin{aligned} Pr(Free = 1) &= \frac{1}{1 + exp(-(-12.4 + 1.5\log(GDP/capita) - 0.68A - 4.7I - 0.75L - 1.8O + 2.02W))} \\ &= \frac{1}{1 + exp(-(-12.4 + 1.5\log(400) - 0.68(0) - 4.7(0) - 0.75(0) - 1.8(0) + 2.02(0)))} \\ &= \frac{1}{1 + exp(-(-12.4 + 1.5 * 5.99))} \\ &= 0.032 \end{aligned}$$

Now, let's try it with the highest GDP/capita:

$$\begin{aligned} Pr(Free = 1) &= \frac{1}{1 + exp(-(-12.4 + 1.5\log(GDP/capita) - 0.68A - 4.7I - 0.75L - 1.8O + 2.02W))} \\ &= \frac{1}{1 + exp(-(-12.4 + 1.5\log(54285) - 0.68(0) - 4.7(0) - 0.75(0) - 1.8(0) + 2.02(0)))} \\ &= \frac{1}{1 + exp(-(-12.4 + 1.5 * 10.9))} \\ &= 0.981 \end{aligned}$$

So, for other countries, a maximal change (i.e., the biggest possible change) in GDP/capita, would make a country go from a probability of being free of 0.03 to about 0.98. We could do this for each of the other civilizations. I'll do the West, and you guys can each choose one to do:

$$
\begin{aligned}
Pr(Free = 1) &= \frac{1}{1 + exp(-(-12.4 + 1.5\log(GDP/capita) - 0.68A - 4.7I - 0.75L - 1.8O + 2.02W))} \\
&= \frac{1}{1 + exp(-(-12.4 + 1.5\log(400) - 0.68(0) - 4.7(0) - 0.75(0) - 1.8(0) + 2.02(1)))} \\
&= \frac{1}{1 + exp(-(-12.4 + 1.5 * 5.99 + 2.02))} \\
&= 0.20 \\
Pr(Free = 1) &= \frac{1}{1 + exp(-(-12.4 + 1.5\log(GDP/capita) - 0.68A - 4.7I - 0.75L - 1.8O + 2.02W))} \\
&= \frac{1}{1 + exp(-(-12.4 + 1.5\log(54285) - 0.68(0) - 4.7(0) - 0.75(0) - 1.8(0) + 2.02(1)))} \\
&= \frac{1}{1 + exp(-(-12.4 + 1.5 * 10.9 + 2.02))} \\
&= 0.997
\end{aligned}
$$

We see that the change is smaller for the Western countries because they start out at a higher level.

Civilization	Δ Pr(Free)
Others	0.949
African	
Islamic	
Latin American	
Orthodox	
Western	0.797

So, what can we say about the effect of GDP/capita in all of the different civilizations on the probability of being free?