Problem Set 4 Solutions

Note: Text that is preceded by a "." is the Stata code used in the analysis. Text enclosed in "*"s explains what each piece of code is doing. Where relevant, I have pasted the actual Stata output.

Part I

. clear

. delimit;

. set more off

. log using ps4.log

. insheet using "/Users/nlmiller/Desktop/Poli Sci Lab/PS4/NMC_v4_0.csv"

. drop if year!=2007

Examining coding scheme and looking for missing values

. tab milex, missing

. tab irst, missing

. tab tpop, missing

Replacing missing value codes as dots
. replace milex=. if milex==-9

. replace irst=. if irst==-9

Note that all three variables are already measured in thousands. Generating military expenditure and iron and steel production per capita variables

. gen milex_pc=milex/tpop

. gen irst_pc=irst/tpop

Transforming variables to 0 to 1 scale after finding and storing minimum and maximum of each variable

. egen milex_min=min(milex_pc)

. egen milex_max=max(milex_pc)

. replace milex_pc=(milex_pc-milex_min)/(milex_max-milex_min)

. egen irst_min=min(irst_pc)

. egen irst_max=max(irst_pc)

. replace irst_pc=(irst_pc-irst_min)/(irst_max-irst_min)

Checking to make sure the variables are bounded between 0 and 1

. sum milex_pc

. sum irst_pc

Running regression of military spending per capita on iron and steel production per capita

. reg milex_pc irst_pc

```
      Source |       SS       df       MS              Number of obs =     166
-------------+------------------------------           F(  1,   164) =    8.33
       Model |  .226056907     1  .226056907           Prob > F      =  0.0044
    Residual |  4.45236358   164  .027148558           R-squared     =  0.0483
-------------+------------------------------           Adj R-squared =  0.0425
       Total |  4.67842049   165  .028354064           Root MSE      =  .16477

------------------------------------------------------------------------------
    milex_pc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     irst_pc |   .4295162   .1488484     2.89   0.004       .13561    .7234225
       _cons |   .0790782   .0134548     5.88   0.000     .0525113    .1056451
------------------------------------------------------------------------------
```

The coefficient tells us that moving from the minimum to the maximum in iron and steel production per capita is associated with a change in military expenditures per capita that is equivalent to 42.95% of that variable's range.

Part II

Bivariate regression coefficients

$$\beta_1 = \frac{cov(x,y)}{var(x)}$$

$$\beta_{economygood} = \frac{1.06552}{1.26169} = .8445$$

$$\beta_{dem7} = \frac{2.27577}{5.00448} = .4547$$

Multivariate regression coefficients

$$Y = \beta_0 + \beta_1(economygood) + \beta_2(dem7) + \epsilon_i$$

$$\beta_1 = \frac{cov(x_1, y)}{var(x_1)} - \beta_2 \frac{cov(x_1, x_2)}{var(x_1)}$$

$$\beta_2 = \frac{cov(x_2, y)}{var(x_2)} - \beta_1 \frac{cov(x_1, x_2)}{var(x_2)}$$

$$\beta_1 = \frac{1.06552}{1.26169} - \beta_2 \frac{1.63027}{1.26269} = .844 - 1.29\beta_2$$

$$\beta_2 = \frac{2.275}{5.004} - \beta_1 \frac{1.63027}{5.004} = .454 - .326\beta_1$$

Substituting for β_1

$$\beta_2 = .454 - .326(.844 - 1.29\beta_2)$$

$$\beta_2 = .454 - .275 + .4205\,\beta_2$$

$$.5795\,\beta_2 = .179$$

$$\beta_2 = .3088$$

Substituting for β_2

β_1= .844-1.29(.3088)

β_1= .446

The coefficients do change. For economygood, the multivariate coefficient is .446 whereas the bivariate is .8445. For dem7, the multivariate coefficient is .3088 while the bivariate coefficient is .4547. The reason that both coefficients decline is that there is a positive correlation between the two variables (and both positively affect the dependent variable). Therefore, both bivariate coefficients overstate the effect because they are partially capturing the effect of the other variable that is excluded from the model.

Variable	**Total**	**Direct**	**Indirect**
economygood	.8445	.446	.3985
dem7	.4547	.3088	.1459

The coefficient on economy01 tells us that moving from the minimum to the maximum in views of the economy (thinking the economy has gotten much worse to thinking the economy has gotten much better) is associated with increase corresponding to 59.1 percentage points of the range of obama_approve, holding dem701 constant. The coefficient on dem701 suggests that moving from being a strong Republican to a strong Democrat is associated with an increase that is equal to 62.04 percentage points of the range of obama_approve, holding economy01 constant. The constant tells us that when both economy01 and dem701 are at their minimums, the expected value of obama01 is -.149 (a meaningless value in this case given that it is coded on a 0 to 1 scale).

Part III

. clear

. use "/Users/nlmiller/Desktop/Poli Sci Lab/PS4/section5.dta"

Regressing Obama vote share on nonwhite population percentage

. reg obamapct nonwhitepct

```
      Source |       SS       df       MS              Number of obs =    3110
-------------+------------------------------           F(  1,  3108) =  527.56
       Model |  102256.033     1  102256.033           Prob > F      =  0.0000
    Residual |  602415.242  3108  193.827298           R-squared     =  0.1451
-------------+------------------------------           Adj R-squared =  0.1448
       Total |  704671.276  3109  226.655283           Root MSE      =  13.922

------------------------------------------------------------------------------
    obamapct |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 nonwhitepct |   .3046643   .0132643    22.97   0.000     .2786566     .330672
       _cons |   33.92938   .3402931    99.71   0.000     33.26216    34.59661
------------------------------------------------------------------------------
```

The coefficient on nonwhitepct implies that a 1 percentage point increase in the nonwhite population is associated with a .305 percentage point increase in the vote for Obama. The constant tells us that a county with zero percent nonwhite population would be expected to give Obama 33.9% of the vote.

Generating interaction term between covered counties and nonwhite population percentage

. gen covered_nonwhite=covered*nonwhitepct

Regressing Obama vote share on the covered dummy variable, nonwhite population percentage, and the interaction between the two

. reg obamapct nonwhitepct covered covered_nonwhite

```
      Source |       SS       df       MS              Number of obs =    3110
-------------+------------------------------           F(  3,  3106) =  360.46
       Model |  181979.622     3   60659.874           Prob > F      =  0.0000
    Residual |  522691.654  3106  168.284499           R-squared     =  0.2582
-------------+------------------------------           Adj R-squared =  0.2575
       Total |  704671.276  3109  226.655283           Root MSE      =  12.972

------------------------------------------------------------------------------
    obamapct |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 nonwhitepct |   .3781311   .0212355    17.81   0.000      .336494    .4197681
     covered |  -19.26641   .9957575   -19.35   0.000    -21.21882     -17.314
covered_no~e |   .2481439   .0310677     7.99   0.000     .1872287    .3090591
       _cons |   35.44979    .352135   100.67   0.000     34.75935    36.14023
------------------------------------------------------------------------------
```

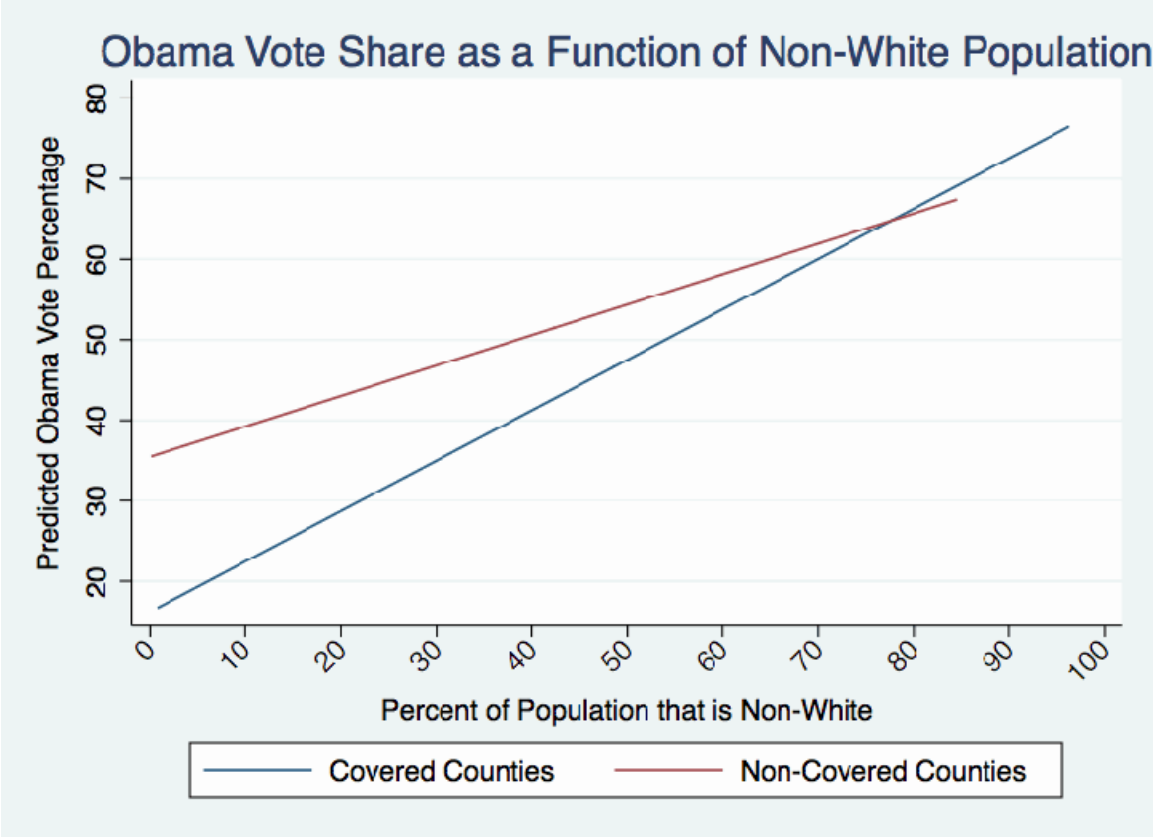
The constant tells us that in non-covered countries with zero percent nonwhite population, the expected vote share for Obama is 35.45%. The nonwhitepct coefficient suggests that in noncovered counties, a one percentage point increase in the nonwhite population is associated with a .378 percentage point increase in the vote for Obama. The coefficient on covered tells us the vertical distance between the intercept for covered and noncovered counties: in other words, covered counties with zero percent nonwhite population are expected to have an Obama vote share that is 19.26 fewer percentage points lower. Finally, the covered_nonwhite coefficient means that in covered counties, a one percentage point increase in the nonwhite population is associated with an *additional* .248 percentage point increase in the Obama vote (on top of the .378 for noncovered counties). Substantively, this suggests that covered counties do indeed vote in a more racially polarized manner: the nonwhite population percentage has a stronger relationship with Obama vote share in covered countries than in noncovered counties.

Storing fitted values

. predict fitted

Plotting fitted values in covered and noncovered counties, customizing axes, titles, and labeling legend to distinguish between the two lines

. twoway (lfit fitted nonwhite if covered==1) (lfit fitted nonwhite if covered==0), title(Obama Vote Share as a Function of Non-White Population) xlabel(0 (10) 100, angle(45)) ylabel (20 (10) 80) ytitle("Predicted Obama Vote Percentage" " ") xtitle("Percent of Population that is Non-White") legend(label(1 "Covered Counties") label(2 "Non-Covered Counties"))


Obama Vote Share as a Function of Non-White Population
Predicted Obama Vote Percentage
20
30
40
50
60
70
80
0
10
20
30
40
50
60
70
80
90
100
Percent of Population that is Non-White
Covered Counties
Non-Covered Counties