Mega HW #3

ECNS 561 (Fall 2017)

Due: 11/7/2017

1.) Work the following Chapter 2 problems from Wooldridge: 3, 5, 6, 8, 10, 12

2.) Work the following Chapter 3 problems from Wooldridge: 3, 5, 8, 11

3.) Change in the sum of squares. Using matrix algebra notation, suppose that **b** is the least squares coefficient vector in the regression of **y** on **X** and that **c** is any other K by 1 vector. Prove that the difference in the two sums of squared residuals is

$$(\mathbf{y} - \mathbf{Xc})'(\mathbf{y} - \mathbf{Xc}) - (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb}) = (\mathbf{c} - \mathbf{b})'\mathbf{X}'\mathbf{X}(\mathbf{c} - \mathbf{b}).$$

Also, prove that this difference is positive.

4.) Linear transformations of the data. Using matrix algebra notation, consider the least squares regression of **y** on K variables (with a constant) **X**. Consider an alternative set of regressors **Z** = **XP**, where **P** is a nonsingular matrix. Thus, each column of **Z** is a mixture of some of the columns of **X**. Prove that the residual vectors in the regressions of **y** on **X** and **y** on **Z** are identical. What relevance does this have to the question of changing the fit of a regression by changing the units of measurement of the independent variables?

5.) In the December, 1969, *American Economic Review* (pp. 886-896), Nathaniel Leff reports the following least squares regression results for a cross-sectional study of the effect of age composition on savings in 74 countries in 1964:

(1) $\ln(S/Y) = 7.3439 + 0.1596{*}\ln(Y/N) + 0.0254{*}\ln G - 1.3520{*}\ln D_1 - 0.3990{*}\ln D_2$

(2) $\ln(S/N) = 8.7851 + 1.1486{*}\ln(Y/N) + 0.0265{*}\ln G - 1.3438{*}\ln D_1 - 0.3966{*}\ln D_2$

where ln(·) represents the natural log, S/Y = domestic savings ratio, S/N = per capita savings, Y/N = per capita income, D_1 = percentage of population under 15, D_2 = percentage of population over 64, and G = growth rate of per capita income. Are these results correct? Explain.

6.) a.) Show that the regression R^2 in the regression of Y on X (i.e., a simple one right-hand-side variable regression) is the squared value of the sample correlation between X and Y. That is, show that $R^2 = r^2_{XY}$

where $r_{XY} = S_{XY}/S_X S_Y$, $S_{XY} = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}[(X_i - \bar{X})(Y_i - \bar{Y})]$, $S^2_X = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}(X_i - \bar{X})^2$, and $S^2_Y = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}(Y_i - \bar{Y})^2$

b.) Show that the R^2 from the regression of Y on X is the same as the R^2 from the regression of X on Y. (Hint: make your life easier and use what you have done in part a.) to quickly explain this result)

STATA Exercise

You are required to complete this problem entirely in Mata and turn in your final do file. The relevant data set is posted on the class webpage under the link "KS Crime Data Set for Mega HW #3."

For what follows, let's only focus on data for the 2011.

a.) Suppose we are interested in estimating the following equation

$$Violent_c = \beta_0 + \beta_1 Unemployment_c + \beta_2 Democrat\_GOP_c + \varepsilon_c$$

where $Violent_c$ is the violent crime rate in county c, $Unemployment_c$ is the unemployment rate in county c, and $Democrat\_GOP_c$ is the ratio of Democratic to GOP votes in county c.

Solve for the OLS estimators. Interpret the coefficient estimates for β_1 and β_2. Are these coefficient estimates meaningful in size?

b.) Compute the standard errors of the estimators, the t-stats, and p-values. Are $\widehat{\beta_1}$ and $\widehat{\beta_2}$ statistically significant at the 5% level? at the 1% level?

c.) Consider the auxiliary regression of

$$Unemployment_c = \alpha_0 + \alpha_1 Democrat\_GOP_c + v_c.$$

Compute the residuals from this auxiliary regression and then regress $Violent_c$ on these residuals. How does the coefficient estimate from the regression of $Violent_c$ on the residuals from the auxiliary equation compare to $\widehat{\beta_1}$ that was computed in part a.). Discuss this result and the underlying intuition.

d.) Add the measure of real income per capita as another control variable to our regression model shown above in part a.). What happens to our coefficient estimate on *Unemployment* when we now control for real income per capita in county c? Is the relationship between *Violent* and *Unemployment* still statistically significant? What is the tradeoff we face when we control for **both** the unemployment rate and real income per capita?