This assignment builds on Assignment #3. If you have not already done assignment #3, you must do it first in order to build on it for this assignment.

1. 2 pts: Present your regression estimates for the replication part of Assignment 3
   a. (That is, run regressions like those in Table 2 of Pendakur and Pendakur 2010, corresponding to regressions controlling for personal characteristics only in Vancouver in 2005 (2006 microdata are about 2005 earnings). Be sure that you get the dependent variable correct. Report the output for the visible minority and Aboriginal coefficients only.)
      i. 2 pts for getting all the P&P regressors and use the logwage as the regressand. 1.5 pt if they leave out HHSIZE (or anything else).
2. 8 pts: Endogeneity NOTE THAT THE STUDENT'S REGRESSION NUMBERS SHOULD BE USED TO EVALUATE.
   a. Control for field of study. Report the resulting regression estimates.
      i. 1 pt for reporting the estimates IF they are different from those in 1. controlling requires dummying out the field of study variable.
   b. What do you do with people who do not report a field of study? (Why don't they report a field of study?)
      i. 1 pt for dealing with the fact that CIP is collinear with HDGREE even when one CIP is dropped. two ways to deal with it: drop CIP=none in addition to one of the field; or drop all persons with HDGREE=no postsec.
   c. What is the meaning of the new coefficients in your regression?
      i. 1 pt for saying that these are marginal effects which hold constant field of study as well as the other covariates
   d. Does inclusion of field of study change your estimates for visible minority or Aboriginal coefficients?
      i. 1 pt for noticing that it does.
   e. Test whether or not visible minority or Aboriginal men or women have lower log earnings than white men or women, respectively.
      i. 2 pts total: 1/2 pt for each fully correct test given the student's estimates and std errs. Since the word "lower" is used, this is a one-tailed test.
   f. If it changes the coefficients, does this mean that the estimates Pendakur and Pendakur (2010) present are biased?
      i. 1/2 pt for "yes", another 1/2 pt for a reason: if the object of interest in P&P was to hold everything INCLUDING THE ERROR TERM constant, then P&P presented biased estimates.
   g. Should you control for field of study?
      i. 1 pt for any argument about what SHOULD be held constant.
3. 9 pts--Heteroskedasticity
   a. Is the variance of the log of earnings smaller different between Aboriginals, Visible Minorities and Whites?
      i. 1 pt for showing that variance of log(WAGES) for these 3 groups (either by sex or not).
   b. If so, does this affect the interpretation of the OLS reported standard errors? How so?
      i. 1/2 pt for: heteroskedasticity is about the variance of the error, not the variance of Y, so it does NOT directly affect the interpretation.
      ii. 1/2 pt for: but, if the variance of Y is big for some groups because the variance of the errors is big for them, then there is heteroskedasticity: V(epsilon) depends on X. Then, OLS is inefficient.
   c. Run the regression (including field of study) reporting White hetero-robust standard errors.
      i. 1 pt for reporting White se's.
   d. Redo the hypothesis tests of 2e. Do the results of the tests change? Why or why not?
      i. 1/2 pt for each test done correctly: do not re-penalise for not noticing that it is one-sided; they should correctly spot changed test results. no parameter estimate is different, only the std errs.
   e. If so, how might this induce inefficiency in the OLS estimator?
      i. 0 points. this question is out of place. it should have appeared above d. it is the same as b.
   f. Can you correct for this using weighted least squares? How does weighted least squares reduce the variance of the estimator?

          i. 1 pt for noting that WLS can correct for heteroskedasticity where you know the variance of epsilon. WLS reduces variance by giving more weight to more informative (lower V(epsilon)) observations.

    g. Estimate the model (including field of study) by weighted least squares. Report the resulting estimates.

          i. 1 pt for WLS using weights that are at all related to V(epsilon) as a function of X. This could be as little as using weights that differ only for the 3 employment equity groups.

    h. Are the estimates different?

          i. 1 pt for showing that the estimates AND std errs are different.

    i. Do the hypothesis tests in 3d change? Why or why not?

          i. 1 pt for the 4 hypothesis tests. do not repenalize for not noticing the one-sidedness.

4. Specification

    a. Instead of regressing log earnings on regressors, regress earnings on regressors (including field of study). Report the regression results.

          i. 1 pt for a regression that uses the same regressors as in 2 but with WAGES as the regressand.

    b. What is the difference in the interpretation of the coefficients in this regression versus that you reported in 2a?

          i. 1 pt for stating that it is a different partial derivative: dWAGES/dX rather than dlnWAGES/dX

    c. Are the coefficients different than in 2a? Are hypothesis tests different than in 2e?

          i. 1/2 pt for each correctly done hypothesis test. do not repenalize for one-sidedness.

    d. No, run the regression reporting White hetero-robust standard errors. Report the results. Do the results of the hypothesis tests change?

          i. 1/2 point for getting White se's (with same coefs), 1/2 pt (total) for the 4 tests.

    e. Run regressions like those in Table 2 of Pendakur and Pendakur 2010, corresponding to regressions controlling for personal **and work** characteristics in Vancouver in 2005. Do not control for field of study. Use regular OLS standard errors (not White hetero-robust ones). Report the output for the visible minority and Aboriginal coefficients only.

          i. 1 pt for getting the regressors right, regardless of the regressand.

    f. How do you interpret the difference between these coefficients and those you reported in 1a?

          i. the question is about the difference between coefs, but they should get 1/2 pt for correctly stating that the coef in the new regression gives the partial derivative (marginal effect) that is relevant for what they did (derivative of lnWAGES or WAGES with respect to group, holding personal and work characteristics constant.)

          ii. 1 pt for stating that the difference between the coefs can be interpreted as the product of the effect of work characteristics on wages multiplied by the covariance of work characteristics with group status. Any similar explanation will do.

    g. Should you control for personal and work characteristics? Or just personal characteristics?

          i. 1 pt for any coherent argument based on what the desired partial derivative is.