ECON 836 Midterm 2016

Each of the eight questions is worth 4 points. You have 2 hours. No calculators, I-devices, computers, and phones. No open books, and everything must be on the floor.

Good luck!

1) Snappers:
   a) [Study Question 6] Why are residuals in regressions mean-zero?
   b) [Study Question 8] How does use of the White hetero-robust covariance matrix estimator affect the estimated coefficients, compared to the OLS estimates?
   c) Why is the Maximum likelihood estimator for the linear model with normal errors equal to the OLS estimator for the linear model with exogenous regressors? You may answer with a proof if you like.
   d) In a panel model with regressors, time effects and unit effects, under what conditions does the OLS regression of the dependent variable on regressors and time effects (but not unit effects) deliver unbiased estimates of the parameters multiplying the regressors.

2) [Study Question 1, modified] Pendakur and Pendakur (2011) control for personal characteristics X, but do not control for job characteristics Z in some of their regressions. If $X=[V\ W]$ where W represents variables of interest (in this case, Aboriginal status) and V represents variables that are controlled for, but are not of direct interest, (in this case, age, education and so on) does it matter if:
   a) V and W are correlated? Can you give an example of this kind of correlation? How does it affect your interpretation of the estimated coefficient on W?
   b) $E[ZZ']$ depends on W? Can you give an example of this kind of dependence? How does it affect your interpretation of the estimated coefficient on W?

3) [Study Question 3, modified] Allen, Pendakur and Suen (2005) estimate a panel model with the standard deviation of the log of age at first marriage on the LHS and no-fault status and state and year dummies on the RHS.
   a) They do not include any information about the population of the state in the model. Likewise, there is not information on education levels in the state. Under what conditions does it not matter? Are these plausible conditions?
   b) Suppose they wanted to implement the random effects FGLS estimator for this model. How would they do it?
   c) How do you think the variance of this estimate would compare to that of the fixed effects estimator that they used?
   d) Why do you think they didn't use the random effects FGLS estimator?

4) [Study Question 8, modified] If you have heteroskedasticity of **unknown** form,
   a) Why is it not possible to use FGLS?
   b) How does the White hetero-robust covariance matrix estimator get around this problem?
   c) How does use of the White covariance matrix affect the estimated coefficients, compared to the OLS estimates?
   d) Is the White covariance matrix valid at all sample sizes, or only asymptotically? Why?

5) Consider the following code and output from a log-wage regression using 2001 Census data on British Columbia residents.

```
#delimit;
generate insamp=pobp<11&agep<65&agep>24&cowp==1&hlosp~=.&
wagesp>0&provp==59;
generate logwage=log(wages);
generate alone=unitsp==1;
recode agep (25/29=1) (30/34=2) (35/39=3) (40/44=4) (45/49=5) (50/54=6) (55/59=7)
        (60/64=8) (else=0), gen(agegp);
generate vismin=visminp<5;
generate aborig=abethncp<3;
replace vismin=0 if aborig==1;
generate white=(vismin==0)&(aborig==0);
xi: regress logwage i.agegp i.hlosp i.marsthp i.cmap alone unitsp i.olnp vismin
aborig if(insamp==1&sexp==2);
```

the Stata output is

```
i.agegp           _Iagegp_0-8         (naturally coded; _Iagegp_0 omitted)
i.hlosp           _Ihlosp_1-14        (naturally coded; _Ihlosp_1 omitted)
i.marsthp         _Imarsthp_1-5       (naturally coded; _Imarsthp_1 omitted)
i.cmap            _Icmap_933-935      (naturally coded; _Icmap_933 omitted)
i.olnp            _Iolnp_1-4          (naturally coded; _Iolnp_1 omitted)

      Source |       SS       df       MS              Number of obs =    8758
-------------+------------------------------           F( 32,  8725) =   39.91
       Model |  1079.87668    32  33.7461464           Prob > F      =  0.0000
    Residual |  7377.23906  8725  .845528832           R-squared     =  0.1277
-------------+------------------------------           Adj R-squared =  0.1245
       Total |  8457.11574  8757   .96575491           Root MSE      =  .91953
------------------------------------------------------------------------------
     logwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   _Iagegp_1 |  -.4752942   .0460058   -10.33   0.000    -.5654763    -.385112
   _Iagegp_2 |  -.2172347   .0430687    -5.04   0.000    -.3016596   -.1328099
   _Iagegp_3 |  -.1215169   .0429977    -2.83   0.005    -.2058026   -.0372312
   _Iagegp_4 |  -.0357658   .0426757    -0.84   0.402    -.1194202    .0478887
   _Iagegp_5 |   .0141983   .0433798     0.33   0.743    -.0708363     .099233
   _Iagegp_6 |    .007959   .0440889     0.18   0.857    -.0784657    .0943837
   _Iagegp_7 |  (dropped)
   _Iagegp_8 |  -.3125657    .055457    -5.64   0.000    -.4212746   -.2038568
   _Ihlosp_2 |   .4623549   .2465691     1.88   0.061    -.0209787    .9456885
   _Ihlosp_3 |   .4276838   .2295029     1.86   0.062     -.022196    .8775636
   _Ihlosp_4 |   .6117929   .2295909     2.66   0.008     .1617405    1.061845
   _Ihlosp_5 |   .6476612   .2331167     2.78   0.005     .1906974    1.104625
   _Ihlosp_6 |   .6004473   .2311027     2.60   0.009     .1474316    1.053463
   _Ihlosp_7 |   .6112917   .2296168     2.66   0.008     .1611886    1.061395
   _Ihlosp_8 |   .7055086   .2293639     3.08   0.002     .2559013    1.155116
   _Ihlosp_9 |   .6810592   .2317898     2.94   0.003     .2266965    1.135422
  _Ihlosp_10 |   .7282987   .2303611     3.16   0.002     .2767367    1.179861
  _Ihlosp_11 |   .9076078   .2292791     3.96   0.000     .4581666    1.357049
  _Ihlosp_12 |   .7355061   .2374668     3.10   0.002     .2700152    1.200997
  _Ihlosp_13 |   .9562865   .2323421     4.12   0.000     .5008413    1.411732
  _Ihlosp_14 |   1.060794   .2409619     4.40   0.000     .5884518    1.533136
 _Imarsthp_2 |   .3709433   .0462105     8.03   0.000     .2803598    .4615268
 _Imarsthp_3 |   .1048837   .0677711     1.55   0.122    -.0279636    .2377311
 _Imarsthp_4 |  -.1250447   .0465665    -2.69   0.007     -.216326   -.0337634
 _Imarsthp_5 |  -.0425507   .1493504    -0.28   0.776    -.3353126    .2502113
  _Icmap_935 |  -.1541335   .0268411    -5.74   0.000    -.2067484   -.1015185
       alone |   .2102727   .0384597     5.47   0.000     .1348827    .2856627
      unitsp |   .0239406   .0092155     2.60   0.009     .0058761    .0420051
    _Iolnp_2 |  -1.468731    .651815    -2.25   0.024    -2.746442   -.1910201
    _Iolnp_3 |  -.0798289   .0347452    -2.30   0.022    -.1479377   -.0117202
    _Iolnp_4 |  -1.053147   .3835622    -2.75   0.006    -1.805019   -.3012743
      vismin |  -.0912763   .0445894    -2.05   0.041    -.1786821   -.0038705
      aborig |  -.1962926    .053485    -3.67   0.000    -.3011358   -.0914494
       _cons |   9.651174    .235676    40.95   0.000     9.189194    10.11315
------------------------------------------------------------------------------
```

a) Why is `_Iagegp_7` dropped?
b) Why is `white` not a regressor in the regression?
c) How is it that so many coefficients are significant, and yet R-squared is only 12%? Does this suggest a problem in the model?
d) The constant is highly significant, with a t-value of 41. Is this surprising? Why or why not?

6) Suppose that I have the nonlinear equation $Y_i = \sum_{k=1}^{K} \left(X_i^k \right)^{\beta_k} + \varepsilon_i,\ E\left[X'\varepsilon \right] = 0_K$. Here, there are K parameters and X is an NxK matrix.
   a) Show how to estimate the parameters of this model by least squares;
   b) Show how to estimate the parameters of this model by the method of moments;

7) Consider the classical linear model with exogenous spherical errors.
   a) Derive the variance of the estimator, using the formula for variance, the formula for the OLS estimator, and the model.
   b) Why is the variance of each estimated coefficient higher when the regressors are highly correlated?
   c) Why doesn't the OLS estimate change if we leave out an uncorrelated regressor?
   d) Derive the variance of the OLS estimator for the case where errors are not spherical, but instead have a non-diagonal covariance matrix Ω.

8) Consider the problem of no-fault divorce and the data of Allen, Pendakur and Suen. Let *state* give the number from 1 to 50 indicating which state the data refer to, and *year* give the year from 1970 to 1989 indicating which year the data refer to. Let *nofault* be a dummy variable indicating whether or not that state-year is no-fault. Let *y* be the dependent variable.
   a) Write Stata code to implement the fixed effects panel regression.
   b) Suppose you wanted to run weighted least squares without using the weight suboption in Stata. Let *number* give the number of observations used to compute the variable *y* in each state-year. Write Stata code to do weighted least squares estimation of their fixed effects model without using the weight option in the regress command.
   c) Suppose *nf1, nf2, and nf3* were indicator variables giving successively stronger definitions of no-fault divorce. These are not mutually exclusive, because if *nf3=1* then both *nf2* and *nf1* equal 1, and if *nf2=1*, then *nf1=1*. Write Stata code that creates a set of mutually exclusive dummy variables, and describe what these dummy variables indicate.
   d) Write Stata code to implement the fixed effects panel regression, but with *fault* as the regressor, where *fault* is an indicator variable equal to 1 when the state-year is fault. How would the estimated coefficient from this regression relate to that from part a) above.