Econ 836 Final Exam

1) [4 points] Let

$$Y = X\beta + \varepsilon,$$
$$X = w + u,$$
$$w \sim N(0, \sigma_w^2 I_N),$$
$$u \sim N(0, \sigma_u^2 I_N),$$
$$\varepsilon \sim N(u\gamma, \sigma_\varepsilon^2 I_N),$$

where X is a just one column. Let $\widehat{\beta}$ denote the OLS estimator, and define residuals e as

$$e = Y - X\widehat{\beta}.$$

Suppose finally that there exists a variable *Z* satisfying

$$Z = w\delta + t,$$
$$t \sim N(0, \sigma_v^2 I_N)$$

Hint: Remember that a multivariate normal vector with a diagonal variance matrix has no correlations across elements of the vector (observations).

a) Derive the bias of $\widehat{\beta}$.
   i) We can write $\varepsilon = u\gamma + v, \quad v \sim N(0, \sigma_\varepsilon^2 I_N)$, so

   $$E\left(\widehat{\beta}\right) - \beta = E\left[(X'X)^{-1} X'\varepsilon\right]$$

   ii) $= E\left[(X'X)^{-1}\right] E[X'\varepsilon] = E\left[(X'X)^{-1}\right] E[(w+u)'(u\gamma + v)] = E\left[(X'X)^{-1}\right]\gamma\sigma_u^2$

   $$= \frac{1}{\sigma_w^2 + \sigma_u^2}\gamma\sigma_u^2$$

   iii) bias is proportional to γ. They should get a point if they can prove it correctly, or even if they leave out stuff in the denominator. If they don't show linearity in γ, they get 0.

b) What is the covariance of *X* and *e*?
   i) 0 by construction of the OLS estimator.

c) Show that *Z* is a valid instrument for *X.*
   i) validity requires exogeneity and nonzero correlation with *X*.
   ii) exogeneity: $E[Z'\varepsilon] = E[(w\delta + t)(u\gamma + v)] = 0$
   iii) relevance: $E[Z'X] = E[(w\delta + t)(w + u)] = \delta\sigma_w^2$

d) Derive the indirect least squares estimator for this model. (The ILS estimator is the moment estimator when rank(*X*)=rank(*Z*).)

   i)
   $$E[Z'\varepsilon] = 0$$
   $$Z'e = 0$$
   $$Z'\left(Y - X\hat{\beta}\right) = 0$$
   $$\hat{\beta} = (Z'X)^{-1} Z'Y$$

2) [4 points] Suppose you have an unbiased estimator of a parameter vector. Suppose also you have estimated the parameter vector with a very large sample, and, found estimates

$$\begin{bmatrix} \widehat{\beta_1} \\ \widehat{\beta_1} \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix},$$

with estimated covariance matrix

$$\widehat{V}\left(\hat{\beta}\right) = \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}.$$

a) What is the expectation of the sum of the two parameters? What is the variance of this sum?
   i) E=-1, V=7
b) Construct a test statistic for the hypothesis that $\beta_1 = \beta_2$. Do you reject the hypothesis?

$$H_0 : R\beta + r = [1 \quad -1]\beta + 0 = 0$$

$$T = \left(R\hat{\beta} + r\right)'\left(RV(\hat{\beta})'R\right)^{-1}\left(R\hat{\beta} + r\right) \sim \chi_1^2$$

$$T = 3\left(RV(\hat{\beta})'R\right)3$$

   i) $$= 3\left([1 \quad -1]\begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right)^{-1} 3 = 3 \sim \chi_1^2$$

$$or,$$

$$3 / \sqrt{\left([1 \quad -1]\begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right)} = \sqrt{3} \sim N(0,1)$$

$$p - value\, 0.85\, ish.$$

   ii) don't reject.

c) Use a Wald Test Statistic to test the hypothesis that $\beta_1 = \beta_2 = 0$. Do you reject the hypothesis?

$$H_0 : R\beta + r = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\beta + \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0$$

$$T = \left(R\hat{\beta} + r\right)'\left(RV(\hat{\beta})'R\right)^{-1}\left(R\hat{\beta} + r\right) \sim \chi_2^2$$

   i) $$T = \begin{bmatrix} 1 & -2 \end{bmatrix}\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)^{-1}\begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -2 \end{bmatrix}\left(\begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}\right)^{-1}\begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -2 \end{bmatrix}\left(\begin{bmatrix} 1/3 & -1/3 \\ -1/3 & 4/3 \end{bmatrix}\right)\begin{bmatrix} 1 \\ -2 \end{bmatrix} = \left(\begin{bmatrix} 5/3 & -9/3 \end{bmatrix}\right)\begin{bmatrix} 2 \\ -1 \end{bmatrix} = 19/3$$

ii) Watch out for ignoring the covariance. The test statistic when the covariance is 0 is equal to 4.25.
iii) reject at 5% level.

d) If the hypothesis in c) is true, then the hypothesis in b) is true. Why are test statistics different?
   i) c implies b, but b doesn't imply c. c has an additional restriction: it says they are equal *and* they are equal to 0.

3) [8 points] Consider the following Stata output, using the data from Assignment 3 (Jacks and Pendakur 2011).

```
. tsset country_num year, yearly
       panel variable:  country_num (strongly balanced)
        time variable:  year, 1870 to 1913
                delta:  1 year
. ********** REGRESSION 1 ************
. regress logaverage freight loggdpsum if (exclude==0 ), robust

Linear regression                                      Number of obs =     734
                                                       F(  2,   731) =  465.06
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4172
                                                       Root MSE      =  .97528

------------------------------------------------------------------------------
             |               Robust
  logaverage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     freight |   .4171854   .1657967     2.52   0.012     .0916909    .7426799
   loggdpsum |   2.591873   .1214102    21.35   0.000     2.353518    2.830227
       _cons |  -13.39943   2.107877    -6.36   0.000    -17.53764   -9.261212
------------------------------------------------------------------------------

. ********** REGRESSION 2 ************
. xtivreg2 logaverage freight loggdpsum if (exclude==0 ), fe robust

FIXED EFFECTS ESTIMATION
------------------------
Number of groups =        21                    Obs per group: min =        12
                                                               avg =      35.0
                                                               max =        44

OLS estimation
--------------

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity

                                                      Number of obs =      734
                                                      F(  2,   711) =   120.63
                                                      Prob > F      =   0.0000
Total (centered) SS     =  67.39455665                Centered R2   =   0.3313
Total (uncentered) SS   =  67.39455665                Uncentered R2 =   0.3313
Residual SS             =  45.06337467                Root MSE      =    .2514

------------------------------------------------------------------------------
             |               Robust
  logaverage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     freight |   .1410819   .0689973     2.04   0.041     .0058497     .276314
   loggdpsum |   .9399335    .095116     9.88   0.000     .7535095    1.126357
------------------------------------------------------------------------------
Included instruments: freight loggdpsum
------------------------------------------------------------------------------

. ********** REGRESSION 3 ************
. xtivreg2 logaverage loggdpsum (freight=L1.(log_steam_tonnage log_sail_tonnage distsailtonnage >
```

```
diststeamtonnage) L2.(log_steam_tonnage log_sail_tonnage distsailtonnage diststeamtonnage)
> logwages logcoal logfish distlogwages distlogcoal distlogfish logaversteam logaversail
> distlogaversteam distlogaversail country_mean country_stddev distcountry_mean
> distcountry_stddev) if exclude==0, fe robust

FIXED EFFECTS ESTIMATION
------------------------
Number of groups =        21                  Obs per group: min =        12
                                                             avg =      33.7
                                                             max =        42

IV (2SLS) estimation
--------------------

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity

                                                      Number of obs =      708
                                                      F(  2,   685) =   129.74
                                                      Prob > F      =   0.0000
Total (centered) SS     =  55.75710693                Centered R2   =   0.3174
Total (uncentered) SS   =  55.75710693                Uncentered R2 =   0.3174
Residual SS             =  38.06144662                Root MSE      =    .2354

------------------------------------------------------------------------------
             |               Robust
  logaverage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     freight |    .052251   .1104144     0.47   0.636    -.1641572    .2686592
   loggdpsum |   .7890528   .1242683     6.35   0.000     .5454915    1.032614
------------------------------------------------------------------------------
Instrumented:         freight
Included instruments: loggdpsum
Excluded instruments: L.log_steam_tonnage L.log_sail_tonnage L.distsailtonnage
                      L.diststeamtonnage L2.log_steam_tonnage
                      L2.log_sail_tonnage L2.distsailtonnage L2.diststeamtonnage
                      logwages logcoal logfish distlogwages distlogcoal
                      distlogfish logaversteam logaversail distlogaversteam
                      distlogaversail country_mean country_stddev
                      distcountry_mean distcountry_stddev
------------------------------------------------------------------------------
```

a) Why is the coefficient on freight in regression 1 (.4171854) so much larger than that in regression 2 (.1410819)?
   i) Regression 2 has country fixed effects. The fixed effects are positively correlated with freight, so excluding them induces positive bias on the freight coefficient.
b) Why is the coefficient on freight in regression 2 so much larger than that in regression 3 (.052251)?
   i) Regression 2 instruments for freight. The model disturbance term is positively correlated with freight. Instrumenting controls for this, and removes the bias.
c) Does the model for regression 2 contain a vector of year dummies? If so, which year is the left-out dummy?
   i) No, they are country dummies.
d) Why is the coefficient in regression 3 insignificant?
   i) The instruments are not very strong, so the first stage does not have much variance in the "freight-hat" regressor. Thus, its standard error is big, and we cannot tell whether or not it is different from 0. Alternatively, we may say that coefficient is zero, and it is insignificant because we did not make a Type I error.
e) Why are there fewer observations in regression 3 than in regression 2?
   i) the instruments include twice lagged variables. You cannot use lagged variables in

the first two periods, so we lose observations from the earliest 2 years of each country.

f) Regression 2 says "`Estimates efficient for homoskedasticity only; Statistics robust to heteroskedasticity` ". Does this mean we should not trust the estimated parameter values if the data are heteroskedastic?
   i) No: the coefficient estimates are unbiased even if the data are heteroskedastic.

g) In regression 3, `loggdpsum` is listed as an included instrument. Does this variable enter the first stage of the 2-stage least squares estimator?
   i) yes.

h) The instruments include `log_steam_tonnage log_sail_tonnage distsailtonnage diststeamtonnage`. These instruments are highly correlated with each other. How does this affect your interpretation of the estimated parameter values?
   i) This does not matter at all. We only need the predicted values from the first stage, not the parameter estimates. So imprecision in the first stage parameter estimates caused by correlation does not hurt.

4) [4 points] If the data come in groups, also known as clusters, one might expect that there could be correlations within clusters but not across clusters. Here, correlations across observations within clusters are of unknown form, but correlations across observations in different clusters are assumed to be zero. For this question, assume there are 2 clusters.
   a) How could one estimate the variance of the OLS estimated parameter vector? (This is called the "clustered" covariance matrix.)
      i) regress Y X, collect e.
      ii) create ee', but with the cross-cluster elements set to 0, call this $\widehat{\Omega}$, create $\widehat{X'\Omega X}$.
      iii) create the sandwich variance estimator: $\widehat{V}\left[\hat{\beta}\right] = (X'X)^{-1}\,\widehat{X'\Omega X}\,(X'X)^{-1}$
   b) What happens if one of the clusters is very big, say, 95% of the observations? Why can't we use this method to model arbitrary cross-observation correlations by saying that there is just one cluster, and using the clustered covariance matrix?
      i) Then, there won't be many zeroes in $\widehat{\Omega}$, and it will look very much like ee'. With just one cluster, it is equal to ee', and you get an X'ee'X, which is 0.
   c) Why is it impossible to implement an FLGS estimator for this problem?
      i) Too many parameters: the number of parameters to estimate in the first stage grows with the square of *N*.
   d) Suppose that within clusters, the covariance across observations is the same for any pair of observations (but it is zero for a pair of observations in different clusters). Now can I implement an FGLS estimator? What is this estimator?
      i) Now I know that within clusters, there are only 2 parameters: the variance, and the covariance across observations. I can get an estimate of these 2 parameters from on OLS first stage.
      ii) This is actually the SUR estimator. (they don't have to notice this.)

5) [2 points] Suppose I have a panel of data for individuals over time.
   a) Can I estimate a panel model just with OLS? Under what conditions is this okay?
      i) If the unit effects are correlated with *X*, then the OLS estimates of the parameters on *X* are biased if the model excludes unit effects. However, if the unit effects are

uncorrelated with *X*, then OLS regression is unbiased. This is okay, but not efficient.

b) Assume that in a panel of individuals over time there are unit effects for each person that you want to control for. However, also assume that you are not really interested in the values of these unit effects. How should you estimate the model if you believe that these unit effects are random and independent of the other RHS variables? How should you estimate the model if you believe that these unit effects are non-stochastic?
   i) if random and independent, then OLS regression is unbiased, but not efficient. The random effects model would apply in this case. The RE model would identify the variance of these unit effects.
   ii) if non-random and independent, then random effects may be inappropriate, because you won't identify the unit effects. However, if one does not *want* the unit effects, then RE is still okay because the unit effects are assumed independent of *X*.

6) [4 points] Suppose I have a time-series of GDP/capita, average education levels, and the dependency ratio for the USA from 1950 to 2008 (the dependency ratio is defined as the number of young and old for every working age person). I am interested in how education affects GDP/capita.
   a) Should I just regress GDP/capita on average education, and the dependency ratio? Why or why not? Be specific.
      i) If time does not belong in the model as a regressor, and if errors are not correlated over time, and if the two regressors are not cointegrated, then OLS regression is okay. If any of these conditions does not hold, then OLS regression could be improved upon. Excluding a relevant time regressor could induce bias; not correcting for correlated errors could induce inefficiency and make OLS reported std errs wrong; including cointegrated regressors induces high variance and, in the limit, infinite variances of estimates.
   b) Should I include time dummies or a linear time trend in the model? Why or why not?
      i) time dummies are not identified---too many regressors.
      ii) a linear time trend could be okay, but it might want a quadratic or even more complicated time effect, because, e.g., the productivity slowdown exogenously slowed growth rates after the 1970s.
   c) Should I suspect that the error term in this regression is integrated? If so, what could I do?
      i) if GDP/capita depends on technology and technology improvements are one-way only (you learn new things and never forget old things), then errors are integrated (because knowledge is integrated). In this case, one could difference the whole model.
   d) Average education is correlated with the dependency ratio (both have grown over time). Is this a problem for estimation?
      i) some correlation is not a big problem unless it induces so much variance in the estimates that you cannot see anything. However, correlation that is integrated (co-integrated regressors) would induce asymptotic non-identification. But, just the fact that both regressors have grown over time is not sufficient for co-integration---they could both contain a linear time trend, for example.

7) [2 points] Consider Dunbar, Lewbel and Pendakur (2011) "Children's Resources in

Collective Households...".

a) What makes their estimation "structural estimation"? What is the structure? Suppose the SAP restriction does not really hold, but we estimate the model given SAP anyways. What do the resulting estimates mean?
   i) the model is 'structural' because we use data to try to recover parameters of a structural model (e.g., eta's). that is, we wrote down a model, it had interesting parameters, and now we go to the data to try to identify those parameters. If SAP is wrong, then the model is wrong, and the estimates in the data do not correspond to any structural parameter, and are consequently meaningless.
b) In their model, budget shares are linear in the log of expenditure. Could they estimate the model by linear regression of budget shares on the log of expenditure and z's, and then use the resulting estimates to construct estimates of the structural parameters?
   i) yes. the slope of budget shares with respect to lny is determined by the product of beta and the relevant eta. For a household with 3 members, there are 3 slopes (one for each person's clothing) and 3 parameters (1 beta, 2 eta's), and therefore slopes can be solved for structural parameters (if the rank condition holds).

Some probabilities for the standard normal distribution, z.
Prob[-1.65<z<1.65]=0.90, Prob[-1.96<z<1.96]=0.95, Prob[-2.56<z<2.56]=0.99

Some probabilities for the chi-square distribution with 1 degree of freedom, x, and 2 degrees of freedom, y:
Prob[x<3.84]=0.95, Prob[x<6.63]=0.99, Prob[y<5.99]=0.95, Prob[x<9.21]=0.99

Mean and Variance Rules for Linear Combinations:
If for some vector *w*: $E(w) = \overline{w}, V(w) = \Sigma$, then for a linear combination *Aw+b*,

we have that $E(Aw+b) = A\overline{w} + b, V(Aw+b) = A\Sigma A'$.

Inverse of a 2x2 matrix:

$$inv\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$