Econ 836 Final Exam

1) [4 points] Let

$$Y = X\beta + \varepsilon,$$
$$\varepsilon \sim N(X\Gamma, \sigma^2 I_N),$$

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 that X is drawn from a multivariate normal distribution with the same mean for every observation:

$$X \sim N(w, \sigma_X^2 I_N),$$

where *w* is an *N*-vector.
Suppose finally that there exists a variable *Z* satisfying

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

a) Derive the bias of $\widehat{\beta}$.
b) What is the covariance of X and e?

2) Suppose you have an unbiased estimator of a parameter vector. Suppose also you have estimated the parameter vector with a very large sample, and, invoking a central limit theorem, have found the sampling distribution of your estimated parameters to be

$$\hat{\beta} = \begin{bmatrix} \widehat{\beta_1} \\ \widehat{\beta_1} \end{bmatrix} \sim N\left(\begin{bmatrix} 1 \\ -2 \end{bmatrix}, \begin{bmatrix} 4 & 1 \\ 1 & 9 \end{bmatrix} \right).$$

a) Test the hypothesis that the two parameters are equal to each other.
b) Use a Wald Test Statistic to test the hypothesis that $\beta_1 = \beta_2 = -1$. Do you reject the hypothesis?
c) Construct a test statistic for the hypothesis that $\beta_1 = \beta_2$. Do you reject the hypothesis?
d) If the hypothesis in b) is true, then the hypothesis in c) is true. Why are test statistics different?

3) 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)?
b) Why is the coefficient on freight in regression 2 so much larger than that in regression 3 (.052251)?
c) Does the model for regression 2 contain a vector of year dummies? If so, which year is the left-out dummy?
d) Why is the coefficient in regression 3 insignificant?
e) Why are there fewer observations in regression 3 than in regression 2?
f) Regression 2 says "`Estimates efficient for homoskedasticity only; Statistics robust to heteroskedasticity` ". Does this mean we should not trust the estimates if the data are homoskedastic?
g) Regression 3 is estimated by 2-stage least squares. Would it be the same if estimated by GMM? Does the skedasticity of the errors matter for this?
h) 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) 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 estimates?