```
//Question 5
clear all
cd "C:/Users/yw3982/Downloads"
use AJR2001.dta

//part a)
reg loggdp risk
**

      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(1, 62)        =     68.17
       Model |  36.2163143         1  36.2163143   Prob > F        =    0.0000
    Residual |  32.9382286        62  .531261752   R-squared       =    0.5237
-------------+----------------------------------   Adj R-squared   =    0.5160
       Total |  69.1545429        63  1.09769116   Root MSE        =    .72888

------------------------------------------------------------------------------
      loggdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        risk |    .516187   .0625186     8.26   0.000     .3912141    .6411599
       _cons |   4.687415    .417441    11.23   0.000     3.852962    5.521867
------------------------------------------------------------------------------

**

reg risk logmort0
**
      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(1, 62)        =     23.34
       Model |  37.1754537         1  37.1754537   Prob > F        =    0.0000
    Residual |  98.7466676        62  1.59268819   R-squared       =    0.2735
-------------+----------------------------------   Adj R-squared   =    0.2618
       Total |  135.922121        63  2.15749399   Root MSE        =     1.262

------------------------------------------------------------------------------
        risk |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    logmort0 |  -.6132892   .1269412    -4.83   0.000    -.8670411   -.3595374
       _cons |   9.365895   .6105941    15.34   0.000     8.145335    10.58646
------------------------------------------------------------------------------



**

ivregress 2sls loggdp (risk= logmort0)
**
Instrumental variables (2SLS) regression          Number of obs   =         64
                                                  Wald chi2(1)    =      36.60
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.1880
                                                  Root MSE        =     .93672

------------------------------------------------------------------------------
      loggdp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        risk |   .9294897   .1536318     6.05   0.000     .6283768    1.230603
       _cons |   1.994296   1.007904     1.98   0.048     .0188405    3.969751
------------------------------------------------------------------------------
Instrumented:  risk
Instruments:   logmort0


**
** Both OLS regression and 2SLS regression have point estimates that are different by 0.01 from the the reported value.



//part b)
**To find homoskedastic and heteroskedastic standard error, I run both robust and
```

```
non-robust regression below.

regress loggdp risk,vce(robust)
**

Linear regression                               Number of obs     =         64
                                                F(1, 62)          =     102.04
                                                Prob > F          =     0.0000
                                                R-squared         =     0.5237
                                                Root MSE          =     .72888

------------------------------------------------------------------------------
             |               Robust
      loggdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        risk |    .516187    .051101    10.10   0.000     .4140376    .6183364
       _cons |   4.687415   .3244741    14.45   0.000       4.0388    5.336029
------------------------------------------------------------------------------




**

regress loggdp risk
**
      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(1, 62)        =     68.17
       Model |  36.2163143         1  36.2163143   Prob > F        =    0.0000
    Residual |  32.9382286        62  .531261752   R-squared       =    0.5237
-------------+----------------------------------   Adj R-squared   =    0.5160
       Total |  69.1545429        63  1.09769116   Root MSE        =    .72888

------------------------------------------------------------------------------
      loggdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        risk |    .516187   .0625186     8.26   0.000     .3912141    .6411599
       _cons |   4.687415    .417441    11.23   0.000     3.852962    5.521867
------------------------------------------------------------------------------

**

regress risk logmort0,vce(robust)
**
Linear regression                               Number of obs     =         64
                                                F(1, 62)          =      16.33
                                                Prob > F          =     0.0001
                                                R-squared         =     0.2735
                                                Root MSE          =      1.262

------------------------------------------------------------------------------
             |               Robust
        risk |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    logmort0 |  -.6132892   .1517849    -4.04   0.000     -.916703   -.3098755
       _cons |   9.365895   .7084137    13.22   0.000     7.949796    10.78199
------------------------------------------------------------------------------

**

regress risk logmort0
**

      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(1, 62)        =     23.34
       Model |  37.1754537         1  37.1754537   Prob > F        =    0.0000
    Residual |  98.7466676        62  1.59268819   R-squared       =    0.2735
-------------+----------------------------------   Adj R-squared   =    0.2618
       Total |  135.922121        63  2.15749399   Root MSE        =     1.262

------------------------------------------------------------------------------
```

```
        risk |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    logmort0 |  -.6132892   .1269412    -4.83   0.000    -.8670411   -.3595374
       _cons |   9.365895   .6105941    15.34   0.000     8.145335    10.58646
------------------------------------------------------------------------------

**
** For 2sls regression, I run three regressions with robust, unadjusted, and cluster respectively.

ivregress 2sls loggdp (risk = logmort0),vce(robust)
**

Instrumental variables (2SLS) regression          Number of obs   =         64
                                                  Wald chi2(1)    =      29.86
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.1880
                                                  Root MSE        =     .93672

------------------------------------------------------------------------------
             |               Robust
      loggdp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        risk |   .9294897   .1700872     5.46   0.000     .5961249    1.262855
       _cons |   1.994296   1.134985     1.76   0.079    -.2302345    4.218826
------------------------------------------------------------------------------
Instrumented:  risk
Instruments:   logmort0

**

ivregress 2sls loggdp (risk = logmort0), vce(cluster risk)
**
Instrumental variables (2SLS) regression          Number of obs   =         64
                                                  Wald chi2(1)    =      30.61
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.1880
                                                  Root MSE        =     .93672

                                    (Std. Err. adjusted for 48 clusters in risk)
------------------------------------------------------------------------------
             |               Robust
      loggdp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        risk |   .9294897   .1680151     5.53   0.000     .6001861    1.258793
       _cons |   1.994296    1.13931     1.75   0.080    -.2387104    4.227302
------------------------------------------------------------------------------
Instrumented:  risk
Instruments:   logmort0

**

ivregress 2sls loggdp (risk = logmort0),vce(unadjusted)
**
Instrumental variables (2SLS) regression          Number of obs   =         64
                                                  Wald chi2(1)    =      36.60
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.1880
                                                  Root MSE        =     .93672

------------------------------------------------------------------------------
      loggdp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        risk |   .9294897   .1536318     6.05   0.000     .6283768    1.230603
       _cons |   1.994296   1.007904     1.98   0.048     .0188405    3.969751
------------------------------------------------------------------------------
Instrumented:  risk
Instruments:   logmort0

**
```

```
** According to above stata output, the homoskedastic standard errors for all kinds of regression are closer to the reported output of the paper. Therefore, the authors use homoskedastic standard errors.

//part c
** First get the predicted risk
regress risk logmort0
**
      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(1, 62)        =     23.34
       Model |  37.1754537         1  37.1754537   Prob > F        =    0.0000
    Residual |  98.7466676        62  1.59268819   R-squared       =    0.2735
-------------+----------------------------------   Adj R-squared   =    0.2618
       Total |  135.922121        63  2.15749399   Root MSE        =     1.262

------------------------------------------------------------------------------
        risk |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    logmort0 |  -.6132892   .1269412    -4.83   0.000    -.8670411   -.3595374
       _cons |   9.365895   .6105941    15.34   0.000     8.145335    10.58646
------------------------------------------------------------------------------


**
predict risk_hat

**Second, get the predicted loggdp based on logmort0
regress loggdp logmort0
**

      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(1, 62)        =     53.77
       Model |  32.1177728         1  32.1177728   Prob > F        =    0.0000
    Residual |  37.0367701        62   .59736726   R-squared       =    0.4644
-------------+----------------------------------   Adj R-squared   =    0.4558
       Total |  69.1545429        63  1.09769116   Root MSE        =     .7729

------------------------------------------------------------------------------
      loggdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    logmort0 |   -.570046   .0777424    -7.33   0.000    -.7254509   -.4146412
       _cons |    10.6998   .3739452    28.61   0.000     9.952292     11.4473
------------------------------------------------------------------------------


**
predict loggdp_hat


** Third, run the regression of predicted loggdp on predicted risk
regress loggdp_hat risk_hat
**


      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(1, 62)        >  99999.00
       Model |   32.117774         1   32.117774   Prob > F        =    0.0000
    Residual |  3.2990e-12        62  5.3210e-14   R-squared       =    1.0000
-------------+----------------------------------   Adj R-squared   =    1.0000
       Total |   32.117774        63  .509805937   Root MSE        =    2.3e-07

------------------------------------------------------------------------------
  loggdp_hat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    risk_hat |   .9294897   3.78e-08  2.5e+07   0.000     .9294897    .9294898
       _cons |   1.994295   2.48e-07  8.0e+06   0.000     1.994295    1.994296
------------------------------------------------------------------------------
**
** The coefficients are the same. However, the standard error is different from the 2sls regression output.


```

```
//part d
** the regression output of first stage and second stage done separately is the following:
regress risk logmort0
**

      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(1, 62)        =     23.34
       Model |  37.1754537         1  37.1754537   Prob > F        =    0.0000
    Residual |  98.7466676        62  1.59268819   R-squared       =    0.2735
-------------+----------------------------------   Adj R-squared   =    0.2618
       Total |  135.922121        63  2.15749399   Root MSE        =     1.262

------------------------------------------------------------------------------
        risk |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    logmort0 |  -.6132892   .1269412    -4.83   0.000    -.8670411   -.3595374
       _cons |   9.365895   .6105941    15.34   0.000     8.145335    10.58646
------------------------------------------------------------------------------


**
predict risk_hat
regress loggdp risk_hat
**
      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(1, 62)        =     53.77
       Model |  32.1177716         1  32.1177716   Prob > F        =    0.0000
    Residual |  37.0367713        62  .597367279   R-squared       =    0.4644
-------------+----------------------------------   Adj R-squared   =    0.4558
       Total |  69.1545429        63  1.09769116   Root MSE        =     .7729

------------------------------------------------------------------------------
      loggdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    risk_hat |   .9294897    .126763     7.33   0.000      .676094    1.182885
       _cons |   1.994296   .8316306     2.40   0.020     .3318896    3.656701
------------------------------------------------------------------------------


**
** The coefficients are the same; however, the standard errors are different from 2sls
regression output.

//part e)
** run the first stage regression and estimate the residual
regress risk logmort0
**
      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(1, 62)        =     23.34
       Model |  37.1754537         1  37.1754537   Prob > F        =    0.0000
    Residual |  98.7466676        62  1.59268819   R-squared       =    0.2735
-------------+----------------------------------   Adj R-squared   =    0.2618
       Total |  135.922121        63  2.15749399   Root MSE        =     1.262

------------------------------------------------------------------------------
        risk |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    logmort0 |  -.6132892   .1269412    -4.83   0.000    -.8670411   -.3595374
       _cons |   9.365895   .6105941    15.34   0.000     8.145335    10.58646
------------------------------------------------------------------------------


**
predict e_hat, residual

** Now run the regression with the residual
regress loggdp risk e_hat
**      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(2, 61)        =     56.67
       Model |  44.9573006         2  22.4786503   Prob > F        =    0.0000
    Residual |  24.1972423        61  .396676104   R-squared       =    0.6501
-------------+----------------------------------   Adj R-squared   =    0.6386
```

```
       Total |  69.1545429        63  1.09769116   Root MSE        =    .62982

------------------------------------------------------------------------------
      loggdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        risk |   .9294897   .1032975     9.00   0.000     .7229335    1.136046
       e_hat |     -.5689   .1211919    -4.69   0.000    -.8112382   -.3265617
       _cons |   1.994296   .6776848     2.94   0.005     .6391811     3.34941
------------------------------------------------------------------------------

**
** The estimated coefficient is the same as the 2SLS estimate. However, the standard error
is different.


//part f)
regress loggdp risk latitude africa
**

      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(3, 60)        =     39.47
       Model |  45.8984283         3  15.2994761   Prob > F        =    0.0000
    Residual |  23.2561146        60  .387601911   R-squared       =    0.6637
-------------+----------------------------------   Adj R-squared   =    0.6469
       Total |  69.1545429        63  1.09769116   Root MSE        =    .62258

------------------------------------------------------------------------------
      loggdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        risk |   .3765228   .0608421     6.19   0.000     .2548205    .4982252
    latitude |   1.382463   .6440401     2.15   0.036     .0941905    2.670735
      africa |  -.7232696   .1712967    -4.22   0.000    -1.065914   -.3806251
       _cons |   5.652234   .4152858    13.61   0.000     4.821539    6.482929
------------------------------------------------------------------------------



**

** The p-values for latitude and africa are both smaller than 0.05. Therefore, they are
statistically significant. To be safe, run the F-statistics

test latitude africa
**
 ( 1)  latitude = 0
 ( 2)  africa = 0

       F(  2,    60) =   10.84
            Prob > F =    0.0001

**
di "The critical value is " invFtail(2,60,0.05)
**The critical value is 3.1504113
** The F statistics is larger than the critical value. Therefore, they are also jointly
significant. Now, let me compare the R square before and after adding these two variables.
regress loggdp risk
**
      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(1, 62)        =     68.17
       Model |  36.2163143         1  36.2163143   Prob > F        =    0.0000
    Residual |  32.9382286        62  .531261752   R-squared       =    0.5237
-------------+----------------------------------   Adj R-squared   =    0.5160
       Total |  69.1545429        63  1.09769116   Root MSE        =    .72888

------------------------------------------------------------------------------
      loggdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        risk |    .516187   .0625186     8.26   0.000     .3912141    .6411599
       _cons |   4.687415    .417441    11.23   0.000     3.852962    5.521867
------------------------------------------------------------------------------

```

```
**

** The stata output shows that both R-squared and adjusted R-squared increase by over 0.1 after adding latitude and africa. Therefore, the model with these two control variables is doing a better job at explaining the variation of loggdp. Therefore, latitude and africa are predictive of the level of GDP.

//part g)
** The 2sls regression output with control variables is the following:
ivregress 2sls loggdp (risk= logmort0) latitude africa
**

Instrumental variables (2SLS) regression          Number of obs    =         64
                                                  Wald chi2(3)     =      57.55
                                                  Prob > chi2      =     0.0000
                                                  R-squared        =     0.3922
                                                  Root MSE         =     .81039

------------------------------------------------------------------------------
      loggdp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        risk |    .799968   .2497417     3.20   0.001     .3104831    1.289453
    latitude |  -.0553109   1.161701    -0.05   0.962    -2.332203    2.221581
      africa |  -.3479258   .3062581    -1.14   0.256    -.9481806     .252329
       _cons |    2.99507   1.581523     1.89   0.058    -.1046587    6.094798
------------------------------------------------------------------------------
Instrumented:  risk
Instruments:   latitude africa logmort0

**
** In this case, both africa and latitude have the p values that are larger than 0.05. Therefore, they are not statistically significant.

test latitude africa

 ( 1)  latitude = 0
 ( 2)  africa = 0

           chi2(  2) =    1.61
         Prob > chi2 =    0.4471

** The p value is still larger than 0.05. Therefore, latitude and africa are not jointly significant as well. When we run the OLS regression, these two regressors are both individually and jointly significant. However, when we run the 2sls regression, Neither are individually and jointly significant.

//part h)
** First, I need to generate the mortality
gen mort0 = exp(logmort0)
** the reduced form of risk is presented as:
regress risk mort0
**
      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(1, 62)        =      4.24
       Model |  8.69746685         1  8.69746685   Prob > F        =    0.0437
    Residual |  127.224655        62  2.05201056   R-squared       =    0.0640
-------------+----------------------------------   Adj R-squared   =    0.0489
       Total |  135.922121        63  2.15749399   Root MSE        =    1.4325

------------------------------------------------------------------------------
        risk |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       mort0 |  -.0007862   .0003819    -2.06   0.044    -.0015495   -.0000228
       _cons |   6.709419   .2021894    33.18   0.000     6.305248     7.11359
------------------------------------------------------------------------------


**
** The authors prefer logmort0 over mort0 because the logmort0 is more economically and statistically significant than mort0. The coefficient of logmort0 deviates more from 0 than the coefficient of mort0 does. The p value of logmort0 is 0 which is much smaller than that
```

```
for mort0.

//part i)
gen logmort02 = logmort0^2

regress risk logmort0 logmort02
**

      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(2, 61)        =     18.42
       Model |  51.1839123         2  25.5919562   Prob > F        =    0.0000
    Residual |   84.738209        61  1.38915097   R-squared       =    0.3766
-------------+----------------------------------   Adj R-squared   =    0.3561
       Total |  135.922121        63  2.15749399   Root MSE        =    1.1786

------------------------------------------------------------------------------
        risk |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    logmort0 |  -2.645684   .6508988    -4.06   0.000    -3.947237   -1.344132
   logmort02 |   .2101141    .066166     3.18   0.002      .077807    .3424211
       _cons |   13.94859   1.551696     8.99   0.000     10.84579     17.0514
------------------------------------------------------------------------------

**
** The square of logmort0 is statistically signficant. Therefore, the relationship between
risk and logmort0 is nonlinear. Since the regression incorporates quadratic form, we can
intepret the marginal effect of the logmort0.
** We need to run the regression again using specific quadratic form notation so that stata
can identify it when estimating marginal effect.
regress risk logmort0 c.logmort0#c.logmort0
**
      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(2, 61)        =     18.42
       Model |  51.1839124         2  25.5919562   Prob > F        =    0.0000
    Residual |   84.738209        61  1.38915097   R-squared       =    0.3766
-------------+----------------------------------   Adj R-squared   =    0.3561
       Total |  135.922121        63  2.15749399   Root MSE        =    1.1786

---------------------------------------------------------------------------------------
                 risk |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------------+----------------------------------------------------------------
             logmort0 |  -2.645684   .6508988    -4.06   0.000    -3.947237   -1.344132
                      |
c.logmort0#c.logmort0 |   .2101141    .066166     3.18   0.002      .077807    .3424211
                      |
                _cons |   13.94859   1.551696     8.99   0.000     10.84579     17.0514
---------------------------------------------------------------------------------------

sum logmort0,detail
**
               Original Log Settler Mortality
-------------------------------------------------------------
      Percentiles      Smallest
 1%     2.145931       2.145931
 5%      2.70805       2.145931
10%     2.791165       2.701361       Obs                  64
25%     4.232656        2.70805       Sum of Wgt.          64

50%      4.35863                      Mean           4.646749
                        Largest       Std. Dev.      1.252543
75%     5.480639       6.504288
90%     6.284209       7.293018       Variance       1.568863
95%     6.504288       7.602901       Skewness       .3052336
99%     7.986165       7.986165       Kurtosis       3.171949
** Now take 5 percentile, 25 percentile, 50 percentile, 75 percentile, 95 percentile ,and
99 percentile to estimate their conditional marginal effect.
margins, dydx(logmort0) at(logmort0 = (2.70805 4.232656 4.35863 5.480639 6.504288 7.986165))
**
Conditional marginal effects                    Number of obs     =         64
Model VCE    : OLS
```

```

Expression   : Linear prediction, predict()
dy/dx w.r.t. : logmort0

1._at        : logmort0        =     2.70805

2._at        : logmort0        =    4.232656

3._at        : logmort0        =     4.35863

4._at        : logmort0        =    5.480639

5._at        : logmort0        =    6.504288

6._at        : logmort0        =    7.986165

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
logmort0     |
         _at |
          1  |  -1.507685   .3055837    -4.93   0.000    -2.118738   -.8966331
          2  |  -.8670031   .1429619    -6.06   0.000    -1.152873   -.5811331
          3  |  -.8140652   .1343585    -6.06   0.000    -1.082732   -.5453987
          4  |  -.3425655    .146023    -2.35   0.022    -.6345566   -.0505745
          5  |   .0876007   .2505381     0.35   0.728    -.4133812    .5885825
          6  |    .710327   .4333453     1.64   0.106    -.1562005    1.576855
------------------------------------------------------------------------------
**
** This output suggests that the marginal effect is increasing. At first, risk decreases
when logmort0 rises. Laster, risk rises as logmort0 rises. The curve relating to
relationship between logmort0 and risk should be conves.
** The output of the 2sls regression with risk instrumented by both logmort0 and logmort02
is the following:
ivregress 2sls loggdp (risk = logmort0 logmort02)
**
Instrumental variables (2SLS) regression          Number of obs   =         64
                                                  Wald chi2(1)    =      46.68
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.3948
                                                  Root MSE        =     .80865

------------------------------------------------------------------------------
      loggdp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        risk |   .7722554   .1130303     6.83   0.000       .55072    .9937908
       _cons |   3.018849   .7434204     4.06   0.000     1.561772    4.475926
------------------------------------------------------------------------------
Instrumented:  risk
Instruments:   logmort0 logmort02

**
** The coefficient of risk in the 2sls regression with risk instrumented only by logmort0
is  .9294897 while the risk coefficient falls to .7722554 after risk is instrumented by
both logmort0 and its square. The constant rises from 1.994296 to 3.018849 as the square of
logmort0 is introduced to instrumental variables in the 2sls regression. While there is no
change in p value of risk before and after logmort02 is added to instrumental variable, the
p value of the constant drops from 0.048 to 0.

// part j)
** The result of the endogeneity test is the following:
estat endogeneous
**
  Tests of endogeneity
  Ho: variables are exogenous

  Durbin (score) chi2(1)          =  10.4601  (p = 0.0012)
  Wu-Hausman F(1,61)              =  11.9176  (p = 0.0010)
**
```

```
** Since p value of the Hausman test is smaller than 0.05, we reject the null hypothesis that variables are exogeneous at 5% significance level. Therefore, variables are endogeneous.

// part k)

** Do the first stage F test:
reg risk logmort0 logmort02
**

      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(2, 61)        =     18.42
       Model |  51.1839123         2  25.5919562   Prob > F        =    0.0000
    Residual |   84.738209        61  1.38915097   R-squared       =    0.3766
-------------+----------------------------------   Adj R-squared   =    0.3561
       Total |  135.922121        63  2.15749399   Root MSE        =    1.1786

------------------------------------------------------------------------------
        risk |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    logmort0 |  -2.645684   .6508988    -4.06   0.000    -3.947237   -1.344132
   logmort02 |   .2101141    .066166     3.18   0.002      .077807    .3424211
       _cons |   13.94859   1.551696     8.99   0.000     10.84579     17.0514
------------------------------------------------------------------------------

**
test logmort0 logmort02
**
 ( 1)  logmort0 = 0
 ( 2)  logmort02 = 0

       F(  2,    61) =   18.42
            Prob > F =    0.0000

**
** F-statistics is greater than 10 (18.42>10). Instrumental variables therefore are relevant.

//part l)

** Do the J-test:
ivregress 2sls loggdp (risk = logmort0 logmort02)
**
Instrumental variables (2SLS) regression          Number of obs   =         64
                                                  Wald chi2(1)    =      46.68
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.3948
                                                  Root MSE        =     .80865

------------------------------------------------------------------------------
      loggdp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        risk |   .7722554   .1130303     6.83   0.000       .55072    .9937908
       _cons |   3.018849   .7434204     4.06   0.000     1.561772    4.475926
------------------------------------------------------------------------------
Instrumented:  risk
Instruments:   logmort0 logmort02

**
drop e_hat
predict e_hat, residual
regress e_hat logmort0 logmort02
**
      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(2, 61)        =      2.66
       Model |  3.35810086         2  1.67905043   Prob > F        =    0.0780
    Residual |  38.4926828        61  .631027586   R-squared       =    0.0802
-------------+----------------------------------   Adj R-squared   =    0.0501
       Total |  41.8507836        63  .664298153   Root MSE        =    .79437

------------------------------------------------------------------------------
```

```
       e_hat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    logmort0 |   .7516194   .4386953     1.71   0.092     -.125606    1.628845
   logmort02 |  -.0876735   .0445948    -1.97   0.054    -.1768463    .0014993
       _cons |  -1.464118   1.045818    -1.40   0.167    -3.555361    .6271237
------------------------------------------------------------------------------

**

test logmort0 logmort02
**
 ( 1)  logmort0 = 0
 ( 2)  logmort02 = 0

       F(  2,    61) =    2.66
            Prob > F =    0.0780

**
di "The J-test = 2 * 2.66 = " 2* 2.66
**The J-test = 2 * 2.66 = 5.32

** In this case, l = 2, k =1, so l-k = 1

di "The critical value of chi square at 1 degree of freedom at 10% significance level is " invchi2tail(1,0.01)
**The critical value of chi square at 1 degree of freedom at 10% significance level is 2.7055435
** 5.32 > 2.7055435. we reject the null hypothesis that IVs are exogeneous at 10% significance level.

di "The critial value of chi square at 1 degree of freedom at 5% significance level is " invchi2tail(1,0.05)
**The critial value of chi square at 1 degree of freedom at 5% significace level is 3.8414588
** 5.32 > 3.8414588. We reject the null hypothesis that IVs are exogeneous (all coefficients of IVs are zero) at 5% significance level.

di "The critical value of chi square at 1 degree of freedom at 1% significance level is " invchi2tail(1,0.01)
**The critical value of chi square at 1 degree of freedom at 1% significance level is 6.6348966
** 5.32 < 6.6348966. We cannot reject the null hypothesis that the IVs are exogeneous at 1% significance level.

** In conclusion, the instruments are exogeneous at 1% significance level; however, the instrumental variables are not exogeneous at 5% and 10% significance level.

```