OLS vs WLS vs GLS

Simulated Data

```
. set obs 10
obs was 0, now 10

. set seed 123

. g x1=rnormal(10,100)

. g x2=rnormal(27,440)

. g x3=rnormal(-30,100)

. g u12=100*x3^2+rnormal(0,1)

. g u1=sqrt(u12)

. g y=150+0.1*x1-2.7*x2+u1
```

OLS

```
. reg y x1 x2

      Source |       SS       df       MS              Number of obs =      10
-------------+------------------------------           F(  2,     7) =   25.22
       Model |  23532457.5     2  11766228.8           Prob > F      =  0.0006
    Residual |  3265852.78     7  466550.397           R-squared     =  0.8781
-------------+------------------------------           Adj R-squared =  0.8433
       Total |  26798310.3     9  2977590.03           Root MSE      =  683.04

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .3206024   1.574711     0.20   0.844    -3.402997    4.044202
          x2 |  -3.480704   .5189972    -6.71   0.000    -4.707937   -2.253471
       _cons |   924.4407   239.2862     3.86   0.006     358.6187    1490.263
------------------------------------------------------------------------------

. predict uhat, r

. g uhat2=uhat^2

. g x32=x3^2

. reg uhat2 x32, noconst

      Source |       SS       df       MS              Number of obs =      10
-------------+------------------------------           F(  1,     9) =   71.22
       Model |  4.2430e+12     1  4.2430e+12           Prob > F      =  0.0000
    Residual |  5.3614e+11     9  5.9571e+10           R-squared     =  0.8878
-------------+------------------------------           Adj R-squared =  0.8754
       Total |  4.7791e+12    10  4.7791e+11           Root MSE      =  2.4e+05

------------------------------------------------------------------------------
       uhat2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         x32 |   30.81989   3.651869     8.44   0.000     22.55879    39.08099
------------------------------------------------------------------------------

. predict vhat2, xb

. g vhat=sqrt(vhat2)
```

WLS

```
. vwls y x1 x2, sd(vhat)

Variance-weighted least-squares regression           Number of obs   =        10
Goodness-of-fit chi2(7)    =     6.83                Model chi2(2)   =    416.60
Prob > chi2                =   0.4467                Prob > chi2     =    0.0000
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   -.512651   .2449598    -2.09   0.036    -.9927635   -.0325385
          x2 |  -3.022738   .1489636   -20.29   0.000    -3.314702   -2.730775
       _cons |   591.2014   97.31571     6.08   0.000     400.4662    781.9367
------------------------------------------------------------------------------
```

Matrix Approach

```
. g one=1

. mkmat one x1 x2, matrix(X)

. mkmat y, matrix(Y)
```

OLS

```
. matrix b=inv(X'*X)*(X'*Y)

. matrix list b

b[3,1]
             y
one  924.44072
 x1   .3206024
 x2  -3.480704

. matrix u=Y-X*b

. matrix V=u*u'

. matrix u2=vecdiag(V)

. matrix u2t=u2'

. mkmat x32, matrix(X32)

. matrix b2=inv(X32'*X32)*X32'*u2t

. matrix u2hat=X32*b2

. matrix u2hatT=u2hat'

. matrix Vhat=diag(u2hatT)

. matrix list Vhat

symmetric Vhat[10,10]
            r1         r2         r3         r4         r5         r6         r7         r8         r9        r10
 r1  2063.3953
 r2          0  62724.285
 r3          0          0  281633.33
 r4          0          0          0  27580.177
 r5          0          0          0          0  1991510.4
 r6          0          0          0          0          0  68397.079
 r7          0          0          0          0          0          0  9538.9761
 r8          0          0          0          0          0          0          0  158188.26
 r9          0          0          0          0          0          0          0          0   348182.8
r10          0          0          0          0          0          0          0          0          0  204452.88
```

GLS – WLS

```
. matrix bgls=inv(X'*inv(Vhat)*X)*(X'*inv(Vhat)*Y)

. matrix list bgls

bgls[3,1]
              y
one   591.20145
 x1  -.51265099
 x2  -3.0227385

. vwls y x1 x2, sd(vhat)

Variance-weighted least-squares regression           Number of obs   =        10
Goodness-of-fit chi2(7)    =    6.83                 Model chi2(2)   =    416.60
Prob > chi2                =  0.4467                 Prob > chi2     =    0.0000
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   -.512651   .2449598    -2.09   0.036    -.9927635   -.0325385
          x2 |  -3.022738   .1489636   -20.29   0.000    -3.314702   -2.730775
       _cons |   591.2014   97.31571     6.08   0.000     400.4662    781.9367
------------------------------------------------------------------------------
```

OLS vs GLS

An example model is as follows:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i \qquad (i = 1, \ldots, n)$$

Where: y_i = Salary
x_{2i} = Education
x_{3i} = Begin Salary

The model can be stated in matrix notation as:

$$y = X\beta + \varepsilon$$

To illustrate the example of how to estimate parameter by matrix follow OLS method, the STATA commands are as follows:

```
. g one=1

. set matsize 474

Current memory allocation

                    current                                 memory usage
    settable          value     description                 (1M = 1024k)
    --------------------------------------------------------------------
    set maxvar         5000     max. variables allowed           1.733M
    set memory           10M    max. data space                 10.000M
    set matsize         474     max. RHS vars in models          1.754M
                                                            -----------
                                                                13.486M

. mkmat one educ logsalbegin, matrix(X)

. mkmat logsal, matrix(Y)

. matrix XX=X'*X

. matrix XY=X'*Y

. matrix b=inv(XX)*XY

. matrix list XX

symmetric XX[3,3]
                    one         educ  logsalbegin
        one         474
       educ        6395        90215
logsalbegin   4583.2979    62165.992    44376.649

. matrix list XY

XY[3,1]
                logsal
        one  4909.1198
       educ  66609.446
logsalbegin  47527.044

. matrix list b

b[3,1]
                logsal
        one  1.6469168
       educ  .02312228
logsalbegin  .86850439
```

The result can be compared with the OLS command:

```
. reg logsal educ logsalbegin
```

```
      Source |       SS       df       MS              Number of obs =     474
-------------+------------------------------           F(  2,   471) =  945.42
       Model |  59.7829614     2  29.8914807           Prob > F      =  0.0000
    Residual |  14.8916586   471   .03161711           R-squared     =  0.8006
-------------+------------------------------           Adj R-squared =  0.7997
       Total |    74.67462   473  .157874461           Root MSE      =  .17781

------------------------------------------------------------------------------
      logsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0231223   .0038936     5.94   0.000     .0154712    .0307733
 logsalbegin |   .8685044   .0318346    27.28   0.000     .8059489    .9310599
       _cons |   1.646917    .274598     6.00   0.000     1.107328    2.186505
------------------------------------------------------------------------------

. predict e, residual

. g s=abs(e)

. g logsals=logsal/s

. g ones=one/s

. g educs=educ/s

. g logsalbegins=logsalbegin/s

. mkmat logsals, matrix(Ys)

. mkmat ones educs logsalbegins, matrix(Xs)

. matrix bs=inv(Xs'*Xs)*Xs'*Ys

. g s2=s^2

. mkmat s2, matrix(S2)

. matrix W=diag(S2)

. matrix bgls=inv(X'*inv(W)*X)*(X'*inv(W)*Y)

. matrix list bs

bs[3,1]
                logsals
        ones  1.6508999
       educs  .02321079
logsalbegins  .86796105

. matrix list bgls

bgls[3,1]
                logsal
        one  1.6508989
       educ  .02321075
logsalbegin  .86796124
```