Feb 23, 20 Stata WLS, FGLS with WAGE data

February 23, 2020

- Name: Jikhan Jeong
- This is for understanding GLS, FGLS in breif
- This code is done in jupterlab with stata kernel in HPC environment to show the results of code so it may some magic command which is not in stata
- Entire data and code is from following website; however, I changed and modified it a little bit for better understanding.
- Ref: http://www3.grips.ac.jp/~yamanota/yamanoCourses.htm (lecture, code, data source), Thank you for Prof. Yamanota
- Related Wooldridge's book (254-260 page) :https://www.amazon.com/Introductory-Econometrics-Modern-Approach-Standalone/dp/130527010X/ref=sr_1_3?keywords=Wooldridge&qid=1583
-

0.1 Data: WAGE1.DTA

0.2 # Part 0. Data Preparing

[1]:
```
use "WAGE1.DTA", clear
```

[2]:
```
%head 1
```

```
     +-------------------------------------------------------------------------------------+
     | wage   educ   exper   tenure   nonwhite   female   married   numdep   smsa   northcen |
     |-------------------------------------------------------------------------------------|
  1. |  3.1     11       2        0          0        1         0        2      1          0 |
     +-------------------------------------------------------------------------------------+
```

[3]:
```
sum
```

```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        wage |        526    5.896103    3.693086        .53      24.98
        educ |        526    12.56274    2.769022          0         18
```

```
       exper |        526    17.01711    13.57216          1         51
      tenure |        526    5.104563    7.224462          0         44
    nonwhite |        526    .1026616    .3038053          0          1
-------------+---------------------------------------------------------
      female |        526    .4790875     .500038          0          1
     married |        526     .608365    .4885804          0          1
      numdep |        526    1.043726    1.261891          0          6
        smsa |        526    .7224335    .4482246          0          1
    northcen |        526    .2509506    .4339728          0          1
-------------+---------------------------------------------------------
       south |        526    .3555133    .4791242          0          1
        west |        526    .1692015    .3752867          0          1
    construc |        526    .0456274    .2088743          0          1
     ndurman |        526    .1140684     .318197          0          1
    trcommpu |        526    .0437262      .20468          0          1
-------------+---------------------------------------------------------
       trade |        526    .2870722    .4528262          0          1
    services |        526    .1007605    .3012978          0          1
    profserv |        526    .2585551    .4382574          0          1
     profocc |        526    .3669202    .4824233          0          1
     clerocc |        526    .1673004    .3735991          0          1
-------------+---------------------------------------------------------
     servocc |        526    .1406844    .3480267          0          1
       lwage |        526    1.623268    .5315382  -.6348783   3.218076
     expersq |        526    473.4354    616.0448          1       2601
     tenursq |        526    78.15019    199.4347          0       1936
```

- Dependent: lwage
-

0.3 Independent: educ female exper expersq

[7]:
```
reg lwage    educ female exper expersq
```

```
      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(4, 521)       =     86.69
       Model |  59.2711314         4  14.8177829   Prob > F        =    0.0000
    Residual |    89.05862       521   .17093785   R-squared       =    0.3996
-------------+----------------------------------   Adj R-squared   =    0.3950
       Total |  148.329751       525   .28253286   Root MSE        =    .41345

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0841361   .0069568    12.09   0.000     .0704692    .0978029
      female |  -.3371868   .0363214    -9.28   0.000    -.4085411   -.2658324
       exper |     .03891   .0048235     8.07   0.000      .029434    .0483859
```

```
     expersq |   -.000686   .0001074    -6.39   0.000     -.000897   -.0004751
       _cons |    .390483   .1022096     3.82   0.000     .1896894    .5912767
------------------------------------------------------------------------------
```

0.3.1 Step1: Generate weigted variable

- Create weight = assumes we know weight matrix = 1/educ^0.5 is a weight matrix (strong assumption)
-

0.4 Multiply weight with Y and X

[ ]:
```
gen w=1/(educ)^0.5              # Just pick weights, not optimal; however, we␣
 ↪assumes this weight is known and optimal
gen weighted_lwage=  lwage*w
gen weighted_female= female*w
gen weighted_educ=   educ*w
gen weighted_exper=  exper*w
gen weighted_expersq=   expersq*w
```

[16]:
```
sum w weighted_lwage weighted_female weighted_educ weighted_exper ␣
 ↪weighted_expersq
```

```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
           w |        524    .2878218    .0419019   .2357023   .7071068
weighted_l~e |        524    .4600634    .1391728  -.1832736   .9346225
weighted_f~e |        524    .1372049     .145159          0   .4472136
weighted_e~c |        524    3.528959    .3967792   1.414214    4.24264
weighted_e~r |        524    5.058787    4.558981   .2357023   29.44486
-------------+---------------------------------------------------------
weighted_e~q |        524    144.6707    208.5306   .2357023   1501.688
```

[18]:
```
reg weighted_lwage  weighted_educ weighted_female weighted_exper␣
 ↪weighted_expersq w, noc
```

```
      Source |       SS           df       MS      Number of obs   =       524
-------------+----------------------------------   F(5, 519)       =   1660.16
       Model |  113.916451         5  22.7832901   Prob > F        =    0.0000
    Residual |  7.12253755       519  .013723579   R-squared       =    0.9412
-------------+----------------------------------   Adj R-squared   =    0.9406
       Total |  121.038988       524  .230990435   Root MSE        =    .11715

------------------------------------------------------------------------------
--
  weighted_lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
```

```
Interval]
-----------------+------------------------------------------------------------
--
   weighted_educ |    .080147    .006435    12.45   0.000     .0675051
.0927889
 weighted_female |  -.3503307   .0354369    -9.89   0.000    -.4199482
-.2807133
  weighted_exper |   .0367367   .0045745     8.03   0.000     .0277498
.0457236
weighted_expersq |  -.0006319    .000099    -6.39   0.000    -.0008264
-.0004375
               w |   .4557085   .0912787     4.99   0.000     .2763872
.6350297
------------------------------------------------------------------------------
--
```

- Condition: Cov(Y) = Cov(error) is not identity deistributed, and we don't know what is weight matrix in GLS
- We need to estimate the weight matrx by applying the following process
- 1. Run OLS
- 2. Get Resigual of OLS = e_ols = e
- 3. Generate Log esitimated q = log(e'e)
- 4. OLS : log(e'e) ~ educ female exper expsq
- 5. Get a predicted value of log(e'e) = hat of log estimated q = hat_log_(e'e)
- 6. Get Omega by plut predicted log estimated q into exponential function -> Omega = exp(hat_log_(e'e)
- 7. Creating weight matrix by weight = 1/(Omega)^1/2
- 8. Multiply weight to both side of original OLS equation -> Y = X$beta$ -> $wY \sim wX$beta
- 9. Run regression wY ~ wX*beta
-

0.5 Ref: Wooldridge page (260 page FGLS estimation processes)

1. Run OLS

[21]:
```
reg lwage educ female exper expersq
```

```
Unknown #command

      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(4, 521)       =     86.69
       Model |  59.2711314         4  14.8177829   Prob > F        =    0.0000
    Residual |    89.05862       521   .17093785   R-squared       =    0.3996
```

```
-------------+----------------------------------   Adj R-squared   =    0.3950
       Total |  148.329751       525   .28253286   Root MSE        =    .41345

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0841361   .0069568    12.09   0.000     .0704692    .0978029
      female |  -.3371868   .0363214    -9.28   0.000    -.4085411   -.2658324
       exper |     .03891   .0048235     8.07   0.000      .029434    .0483859
     expersq |   -.000686   .0001074    -6.39   0.000     -.000897   -.0004751
       _cons |    .390483   .1022096     3.82   0.000     .1896894    .5912767
------------------------------------------------------------------------------
```

0.5.1 2. Get Resigual of OLS = e_ols = e

[22]:
```
predict e, residual
```

0.5.2 3. Generate Log esitimated q = log(e'e)

[23]:
```
gen log_estimated_q = ln(e*e)
```

0.5.3 4. OLS : log(e'e) ~ educ female exper expsq

[27]:
```
reg log_estimated_q  educ female exper expersq
```

```
      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(4, 521)       =      5.73
       Model |  119.579085         4  29.8947712   Prob > F        =    0.0002
    Residual |  2720.03141       521  5.22078965   R-squared       =    0.0421
-------------+----------------------------------   Adj R-squared   =    0.0348
       Total |  2839.61049       525  5.40878189   Root MSE        =    2.2849

------------------------------------------------------------------------------
log_estima~q |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0718669   .0384467     1.87   0.062    -.0036627    .1473964
      female |   -.496624   .2007295    -2.47   0.014    -.8909627   -.1022852
       exper |   .0807454   .0266572     3.03   0.003     .0283766    .1331143
     expersq |  -.0014259   .0005934    -2.40   0.017    -.0025917   -.0002601
       _cons |  -4.532606     .56486    -8.02   0.000     -5.64229   -3.422923
------------------------------------------------------------------------------
```

0.5.4 5. Get a predicted value of log(e'e) = hat of log estimated q = hat_log_(e'e)

[28]:
```
predict predicted_log_estimated_q
```

(option xb assumed; fitted values)

[36]:
```
sum predicted_log_estimated_q
```

```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
predicted_~q |        526   -3.168716    .4772522  -4.231242  -2.096585
```

0.5.5 6. Get Omega by plut predicted log estimated q into exponential function -> Omega = exp(hat_log_(e'e)

[29]:
```
gen omega=exp(predicted_log_estimated_q)
```

[35]:
```
sum omega
```

```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       omega |        526    .0469478    .0219954    .0145343    .1228753
```

0.5.6 7. Creating weight matrix by weight matrix = 1/(Omega)^1/2

[37]:
```
gen fgls_weight=1/(omega)^0.5
```

[38]:
```
sum fgls_weight
```

```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 fgls_weight |        526    5.017777    1.220166    2.852776    8.294736
```

0.5.7 8. Multiply weight to both side of original OLS equation -> Y = X*beta* -> *w*Y ~ *w*Xbeta

[ ]:
```
gen fgls_weighted_lwage    =   lwage*fgls_weight
gen fgls_weighted_female   =    female*fgls_weight
gen fgls_weighted_educ     =   educ*fgls_weight
gen fgls_weighted_exper    =   exper*fgls_weight
gen fgls_weighted_expersq  =   expersq*fgls_weight
```

[43]:
```
sum fgls_weighted_lwage  fgls_weighted_educ fgls_weighted_female ␣
↪fgls_weighted_exper  fgls_weighted_expersq fgls_weight
```

```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
fgls_weig~ge |        526    7.778322    2.115953  -4.546649    16.6817
fgls_weigh~c |        526    61.77805    16.69801          0   106.9765
fgls_weig~le |        526    2.761222     2.98256          0   8.294736
fgls_weigh~r |        526    78.31162    67.04885   4.854187   366.1111
fgls_weigh~q |        526    2190.806    3222.909   4.854187   18353.15
-------------+---------------------------------------------------------
 fgls_weight |        526    5.017777    1.220166   2.852776   8.294736
```

0.5.8 9. Estimate Feasible GLS (FGLS) model

- Run regression wY ~ wX*beta_fgls

[46]:
```
reg fgls_weighted_lwage  fgls_weighted_educ fgls_weighted_female ␣
↪fgls_weighted_exper  fgls_weighted_expersq fgls_weight,noc
```

```
      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(5, 521)       =   1555.59
       Model |  32029.3037         5  6405.86073   Prob > F        =    0.0000
    Residual |  2145.46375       521  4.11797265   R-squared       =    0.9372
-------------+----------------------------------   Adj R-squared   =    0.9366
       Total |  34174.7674       526  64.9710407   Root MSE        =    2.0293

------------------------------------------------------------------------------
-------
  fgls_weighted_lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
----------------------+-------------------------------------------------------
-------
   fgls_weighted_educ |   .0751668   .0065543    11.47   0.000     .0622906
.088043
 fgls_weighted_female |  -.2858389   .0354633    -8.06   0.000    -.3555075
-.2161704
  fgls_weighted_exper |   .0389362   .0045178     8.62   0.000      .030061
.0478115
fgls_weighted_expersq |   -.000693   .0001004    -6.90   0.000    -.0008903
-.0004958
          fgls_weight |   .4755142   .0923694     5.15   0.000     .2940518
.6569765
------------------------------------------------------------------------------
-------
```

- GLS (=WLS, the first regression) : (= we assumes we know weight matrix, but it is not correct = **wrong weight matrix**)
- FGLS with estimated weight matrix

- Compared to standard error of beta coefficient = **FGLS's educeation coefficient's standard error is the smallest among others**
- Using eststo package to show the results:
  (command line) cap ssc install estout
  (command line) sysuse auto, clear
  (command line) eststo clear

---

[48]:
```
eststo: qui reg lwage    educ female exper expersq
eststo: qui reg weighted_lwage  weighted_educ weighted_female weighted_exper␣
↪weighted_expersq w, noc
eststo: qui reg fgls_weighted_lwage  fgls_weighted_educ fgls_weighted_female ␣
↪fgls_weighted_exper  fgls_weighted_expersq fgls_weight,noc
```

```
(est1 stored)

(est2 stored)

(est3 stored)
```

[49]:
```
%html
esttab, label title("OLS, WLS, FGLS Table") html
```

```
This front-end or document format cannot display HTML
```