Takashi Yamano
Fall Semester 2009

Lecture Notes on Advanced Econometrics

Lecture 10: GLS, WLS, and FGLS

Generalized Least Square (GLS)

So far, we have been dealing with heteroskedasticity under OLS framework. But if we knew the variance-covariance matrix of the error term, then we can make a heteroskedastic model into a homoskedastic model.

As we defined before

$$E(uu') = \sigma^2 \Omega = \Sigma .$$

Define further that

$$\Omega^{-1} = P'P$$

P is a "n x n" matrix

Pre-multiply P on a regression model

$$Py = PX\beta + Pu$$

or

$$\tilde{y} = \tilde{X}\beta + \tilde{u}$$

In this model, the variance of $\tilde{u}$ is

$$E(\tilde{u}\tilde{u}') = E(Puu'P') = PE(uu')P' = P\sigma^2\Omega P' = \sigma^2 P\Omega P' = \sigma^2 I$$

Note that $P\Omega P' = I$, because define $P\Omega P' = A$, then $P'P\Omega P' = P'A$. By the definition of P, $\Omega^{-1}\Omega P' = P'A$, thus $P' = P'A$. Therefore, ***A*** must be ***I***.

Because $E(\tilde{u}\tilde{u}') = \sigma^2 I$, the model satisfies the assumption of homoskedasticity. Thus, we can estimate the model by the conventional OLS estimation.

Hence,

$$\begin{aligned}\hat{\beta} &= (\tilde{X}'\tilde{X})^{-1}\tilde{X}\tilde{y} \\ &= (X'P'PX)^{-1}X'P'Py \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y\end{aligned}$$

is the efficient estimator of β. This is called the **Generalized Least Square (GLS)** estimator. Note that the GLS estimators are unbiased when $E(\tilde{u} \mid \tilde{X}) = 0$. The variance of GLS estimator is

$$\text{var}(\hat{\text{B}}) = \sigma^2 (\tilde{X}\tilde{X})^{-1} = \sigma^2 (X'\Omega^{-1}X)^{-1}.$$

Note that, under homoskedasticity, i.e., Ω^{-1}=***I***, GLS becomes OLS.

The problem is, as usual, that we don't know $\sigma^2\, \Omega$ *or* Σ. Thus we have to either assume Σ or estimate Σ empirically. An example of the former is Weighted Least Squares Estimation and an example of the later is Feasible GLS (FGLS).

Weighted Least Squares Estimation (WLS)

Consider a general case of heteroskedasticity.

$$\text{Var}(u_i) = \sigma_i^2 = \sigma^2 \omega_i.$$

Then,

$$E(uu') = \sigma^2 \begin{bmatrix} \omega_1 & 0 & 0 \\ 0 & \omega_2 & 0 \\ & & \\ 0 & 0 & \omega_n \end{bmatrix} = \sigma^2\, \Omega, \quad \textit{thus}\ \Omega^{-1} = \begin{bmatrix} \omega_1^{-1} & 0 & 0 \\ 0 & \omega_2^{-1} & 0 \\ & & \\ 0 & 0 & \omega_n^{-1} \end{bmatrix}.$$

Because of $\Omega^{-1} = P'P$, ***P*** is a n x n matrix whose *i*-th diagonal element is $1/\sqrt{\omega_i}$. By pre-multiplying ***P*** on ***y*** and ***X***, we get

$$y_* = Py = \begin{bmatrix} y_1/\sqrt{\omega_1} \\ y_2/\sqrt{\omega_2} \\ \\ y_n/\sqrt{\omega_n} \end{bmatrix} \quad \textit{and} \quad X_* = PX = \begin{bmatrix} 1/\sqrt{\omega_1} & x_{11}/\sqrt{\omega_1} \ \ldots\ x_{1k}/\sqrt{\omega_1} \\ 1/\sqrt{\omega_2} & x_{21}/\sqrt{\omega_2} \ \ldots\ x_{2k}/\sqrt{\omega_2} \\ & \\ 1/\sqrt{\omega_n} & x_{n1}/\sqrt{\omega_n} \ \ldots\ x_{nk}/\sqrt{\omega_n} \end{bmatrix}.$$

The OLS on y_* *and* X_* is called the Weighted Least Squares (WLS) because each variable is weighted by $\sqrt{\omega_i}$. The question is: where can we find ω_i?

Feasible GLS (FGLS)

Instead of assuming the structure of heteroskedasticity, we may estimate the structure of heteroskedasticity from OLS. This method is called Feasible GLS (FGLS). First, we estimate $\hat{\Omega}$ from OLS, and, second, we use $\hat{\Omega}$ *instead of* Ω.

$$\hat{\beta}_{FGLS} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y$$

There are many ways to estimate FGLS. But one flexible approach (discussed in Wooldridge page 277) is to assume that

$$\text{var}(u \mid X) = u^2 = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k)$$

By taking log of the both sides and using $\hat{u}^2$ instead of u^2, we can estimate

$$\log(\hat{u}^2) = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k + e.$$

The predicted value from this model is $\hat{g}_i = \widehat{\log(\hat{u}^2)}$. We then convert it by taking the exponential into $\hat{\omega}_i = \exp(\hat{g}_i) = \exp(\widehat{\log(\hat{u}^2)}) = \hat{\hat{u}}^2$. We now use WLS with weights $1/\hat{\omega}_i$ *or* $1/\hat{\hat{u}}^2$.

Example 1

```
. * Estimate the log-wage model by using WAGE1.dta with WLS
. * Weight is educ

. * Generate weighted varaibles
. gen w=1/(educ)^0.5
. gen wlogwage=logwage*w
. gen wfemale=female*w
. gen weduc=educ*w
. gen wexper=exper*w
. gen wexpsq=expsq*w

. * Estimate weighted least squares (WLS) model

. reg wlogwage weduc wfemale wexper wexpsq 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

------------------------------------------------------------------------------
wlogwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
```

```
   weduc |    .080147    .006435     12.455   0.000       .0675051    .0927889
 wfemale |  -.3503307   .0354369     -9.886   0.000      -.4199482   -.2807133
  wexper |   .0367367   .0045745      8.031   0.000       .0277498    .0457236
  wexpsq |  -.0006319    .000099     -6.385   0.000      -.0008264   -.0004375
       w |   .4557085   .0912787      4.992   0.000       .2763872    .6350297
------------------------------------------------------------------------------
```

End of Example 1

Example 2

```
. * Estimate reg
. reg logwage educ female exper expsq
(Output omitted)
. predict e, residual
. gen logesq=ln(e*e)


. reg logesq educ female exper expsq
(output omitted)
. predict esqhat
(option xb assumed; fitted values)
. gen omega=exp(esqhat)

. * Generate weighted varaibles
. gen w=1/(omega)^0.5
. gen wlogwage=logwage*w
. gen wfemale=female*w
. gen weduc=educ*w
. gen wexper=exper*w
. gen wexpsq=expsq*w

. * Estimate Feasible GLS (FGLS) model
. reg wlogwage weduc wfemale wexper wexpsq w, noc

  Source |       SS       df       MS               Number of obs =     524
---------+------------------------------            F(  5,   519) = 1569.72
   Model |  31164.1981     5  6232.83962            Prob > F      =  0.0000
Residual |  2060.77223   519  3.97065941            R-squared     =  0.9380
---------+------------------------------            Adj R-squared =  0.9374
   Total |  33224.9703   524   63.406432            Root MSE      =  1.9927

------------------------------------------------------------------------------
wlogwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   weduc |   .0828952   .0069779     11.880   0.000       .0691868    .0966035
 wfemale |  -.2914609   .0349884     -8.330   0.000      -.3601971   -.2227246
  wexper |   .0376525    .004497      8.373   0.000       .0288179    .0464872
  wexpsq |  -.0006592   .0001008     -6.540   0.000      -.0008573   -.0004612
       w |   .3848487   .0950576      4.049   0.000       .1981038    .5715936
------------------------------------------------------------------------------
```

End of Example 2