1

Generalized Least Squares (GLS)

Hüseyin Taştan[1]

[1]Department of Economics
Ph.D. Program - Advanced Econometrics
Yildiz Technical University

November 19, 2013

2

Generalized Least Squares (GLS)

- Instead of the classical assumption $\mathsf{E}[\boldsymbol{u}\boldsymbol{u}^\top \mid \boldsymbol{X}] = \sigma^2 \boldsymbol{I}_n$, we assume that

$$\mathsf{E}[\boldsymbol{u}\boldsymbol{u}^\top \mid \boldsymbol{X}] = \boldsymbol{\Omega}, \quad n \times n, p.d.$$

- Note that now the errors can be nonspherical, ie, heteroskedastic and/or serially correlated.
- Is OLS still unbiased, consistent, efficient? Does Gauss-Markov theorem hold?
- If $\boldsymbol{X}$ is exogenous ($\mathsf{E}[\boldsymbol{u}|\boldsymbol{X}] = \boldsymbol{0}$) then under mild conditions OLS is still and unbiased...
- but it is not efficient. The Gauss-Markov Theorem does not hold.
- There are more efficient estimation methods than OLS.

3

Generalized Least Squares (GLS)

$$\mathsf{E}[\boldsymbol{u}\boldsymbol{u}^\top \mid \boldsymbol{X}] = \boldsymbol{\Omega}, \quad n \times n, p.d.$$

- The error covariance matrix is general enough to include heteroskedasticity and serial correlation.
- Heteroskedasticity: the error variance is not constant across observations. Usually arises in cross-sectional data sets. But there are time series models which explicitly consider time-varying variances (e.g., ARCH/GARCH models in financial econometrics)
- Serial correlation: Off-diagonal entries of $\boldsymbol{\Omega}$ are not zero. Errors are correlated across observations. Usually arises in time-series data.
- In panel data models both heteroskedasticity and serial correlation may be present.

4

Generalized Least Squares (GLS)

- If $\boldsymbol{\Omega}$ matrix is known then we can easily define more efficient estimation procedures than OLS.
- Let $\boldsymbol{P}$ be an upper triangular matrix

$$\boldsymbol{\Omega}^{-1} = \boldsymbol{P}^\top \boldsymbol{P}$$

- Using this we can transform the model:

$$\begin{aligned} \boldsymbol{P}\boldsymbol{y} &= \boldsymbol{P}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{P}\boldsymbol{u} \\ \boldsymbol{y}^* &= \boldsymbol{X}^*\boldsymbol{\beta} + \boldsymbol{u}^* \end{aligned}$$

- Since $\boldsymbol{\Omega}$ is positive definite we can always find the matrix $\boldsymbol{P}$ and transform the data.
- Applying OLS to the transformed model produces efficient estimators.

Generalized Least Squares (GLS)

- GLS transformation:

$$\begin{aligned} \boldsymbol{Py} &= \boldsymbol{PX\beta} + \boldsymbol{Pu} \\ \boldsymbol{y}^* &= \boldsymbol{X}^*\boldsymbol{\beta} + \boldsymbol{u}^* \end{aligned}$$

- Note that in this model, the error covariance matrix is

$$\begin{aligned} \mathsf{E}[\boldsymbol{u}^*\boldsymbol{u}^{*\top}] &= \mathsf{E}[(\boldsymbol{Pu})(\boldsymbol{Pu})^\top] \\ &= \mathsf{E}[\boldsymbol{Puu}^\top\boldsymbol{P}^\top] \\ &= \boldsymbol{P\Omega P}^\top = \boldsymbol{P}(\boldsymbol{P}^\top\boldsymbol{P})^{-1}\boldsymbol{P}^\top \\ &= \boldsymbol{PP}^{-1}(\boldsymbol{P}^\top)^{-1}\boldsymbol{P}^\top = \boldsymbol{I}_n \end{aligned}$$

- This implies that the transformed model satisfies classical assumptions. Thus, applying OLS to the transformed model produces efficient estimators.

Generalized Least Squares (GLS)

- The GLS objective function is

$$(\boldsymbol{y} - \boldsymbol{X\beta})^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})$$

- The GLS estimator is

$$\begin{aligned} \hat{\beta}_{GLS} &= (\boldsymbol{X}^\top\boldsymbol{P}^\top\boldsymbol{PX})^{-1}\boldsymbol{X}^\top\boldsymbol{P}^\top\boldsymbol{Py} \\ &= (\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{y} \end{aligned}$$

- The GLS estimator is consistent and asymptotically normal under certain assumptions.
- In particular we need to know the exact structure of the error covariance matrix.
- If the errors are heteroscedastic and there is no serial correlation we can use weighted least squares (WLS) and feasible GLS (FGLS).

GLS and Heteroskedasticity

- Suppose that $\mathsf{E}[\boldsymbol{uu}^\top \mid \boldsymbol{X}] = \sigma^2\boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ is an $n \times n$ diagonal matrix.
- Since off-diagonal entries are assumed to be zero, there is no serial correlation.
- Usually encountered in cross-sectional and panel data sets.
- Need to know the structure of the error covariance matrix. This may be difficult in practice. But there are general procedures we can use (FGLS)
- Also need to test for heteroskedasticity
- If heteroskedasticity is present then all inference procedures (t, F, LM tests, confidence intervals) are invalid.
- If sample size is large enough we can use heteroskedasticity-robust standard errors
- Otherwise, we need to use Feasible GLS methods, i.e., assume a general form for the error structure and estimate the $\boldsymbol{\Omega}$ matrix to transform the model.

Heteroskedasticity

Example

As an example, consider the following model for household consumption:

$$cons_i = \beta_1 + \beta_2 inc_i + \beta_3 size_i + u_i, \quad i = 1, 2, \ldots, n$$

where inc is income and $size$ is household size. Suppose that

$$\mathsf{E}(u_i | inc_i, size_i) = 0, \quad \forall\, i$$

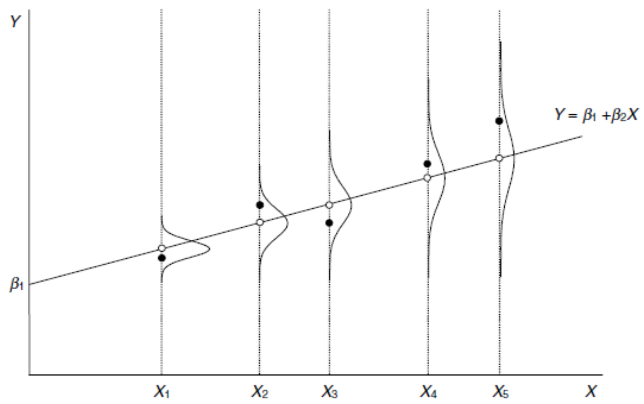
but

$$\mathsf{E}(u_i^2 | inc_i, size_i) = \sigma_i^2 = \sigma^2 inc_i, \quad \sigma^2 > 0, inc_i > 0, \ \forall\, i$$

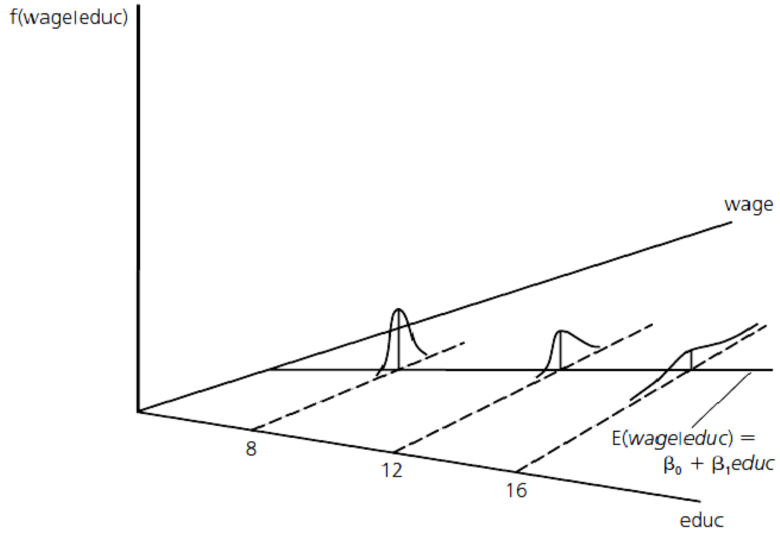
which simply implies that the variability of consumption expenditures is proportional to the level of household income. The model can be written as

$$\mathsf{E}[\boldsymbol{uu}^\top \mid \boldsymbol{X}] = \sigma^2\boldsymbol{\Omega} = \sigma^2 \begin{bmatrix} inc_1 & 0 & 0 & 0 \\ 0 & inc_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & inc_n \end{bmatrix}$$

Heteroscedasticity



Heteroscedasticity: Simple wage equation



GLS and Heteroskedasticity

- OLS estimator

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{u}$$

  is still unbiased (and consistent) under certain assumptions (the crucial one is the exogeneity assumption).
- But inefficient because

$$\text{Var}(\hat{\boldsymbol{\beta}}|\boldsymbol{X}) = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top (\sigma^2 \boldsymbol{\Omega}) \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1}$$

  is larger (in matrix sense) than $\sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}$
- This can also be written as

$$\text{Var}(\hat{\boldsymbol{\beta}}|\boldsymbol{X}) = \frac{\sigma^2}{n} \left(\frac{1}{n} \boldsymbol{X}^\top \boldsymbol{X} \right)^{-1} \left(\frac{1}{n} X^\top \boldsymbol{\Omega} \boldsymbol{X} \right) \left(\frac{1}{n} \boldsymbol{X}^\top \boldsymbol{X} \right)^{-1}$$

- This matrix converges in probability to zero under mild conditions (think of Grenander conditions) which implies that the OLS is (mean-square) consistent.

Testing for Heteroscedasticity

- The null hypothesis is:

$$H_0: \ \text{Var}(u|x_1, x_2, \dots, x_k) = \sigma^2 \qquad \text{constant variance}$$

- Alternative hypothesis is:

$$H_1: \ \text{Var}(u|x_1, x_2, \dots, x_k) \neq \sigma^2 \qquad \text{heteroscedasticity}$$

- Null hypothesis can also be written as

$$H_0: \ \text{E}(u^2|x_1, x_2, \dots, x_k) = \text{E}(u^2) = \sigma^2$$

- This hypothesis says that the variance is not related to x_j. Heteroscedasticity tests detect the presence of this relationship.

13

Testing for Heteroscedasticity

- If the null hypothesis is not correct, the expected value of u^2 can be any function of x_j.
- Assume that this relationship is linear:

$$u^2 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_k x_k + \nu$$

- The null hypothesis of constant variance can be written as

$$H_0 : \ \alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$$

- Under H_0, $\mathrm{E}(u^2|x_1, \ldots, x_k) = \alpha_0$, which is a constant.
- We cannot observe u but we can estimate them. Thus, we can use $\hat{u}$ and carry out an F or LM test.

14

Testing for Heteroscedasticity

- After estimating the model using OLS, we regress squared residuals on all x variables:

$$\hat{u}^2 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_k x_k + \text{hata}$$

And test the joint significance of $x_1, x_2, \ldots, x_k$ using the standard $F(k, n-k-1)$ or $LM = nR^2_{\hat{u}} \sim \chi^2_k$ test procedures.
- If the test statistics are greater than the critical value then we reject the null hypothesis of constant variance in favor of heteroscedasticity.
- The LM version of this test is called **Breusch-Pagan heteroscedasticity test**.

15

Breusch-Pagan Test for Heteroscedasticity

STEPS

1. Estimate the model using OLS as usual, obtain squared residuals $\hat{u}^2$.
2. Run the following test regression (or auxiliary regression):

$$\hat{u}^2 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_k x_k + \text{error}$$

and save the coefficient of determination $R^2_{\hat{u}}$.
3. Using $R^2_{\hat{u}}$ compute either $F \sim F(k, n-k-1)$ or $LM \sim \chi^2_k$ test statistic. (If the test statistic is greater than the critical value reject H_0 : constant variance. This means that there is evidence of heteroscedasticity in the model. Or use p-value: if the p-value is smaller than the chosen significance level, eg 0.05, then we reject the null.)

16

Example: House Prices, hprice1.dta

1st STEP: Estimating the Model using OLS:

$$\widehat{\text{price}} = -\underset{(29.475)}{21.77} + \underset{(0.0006)}{0.0021}\,\text{lotsize} + \underset{(0.013)}{0.123}\,\text{sqrft} + \underset{(9.010)}{13.853}\,\text{bdrms}$$

$$n = 88 \quad R^2 = 0.672$$

2nd STEP: Test regression:

$$\widehat{\text{u}}^2 = -\underset{(3259.5)}{5522.79} + \underset{(0.071)}{0.202}\,\text{lotsize} + \underset{(1.464)}{1.691}\,\text{sqrft} + \underset{(996.38)}{1041.76}\,\text{bdrms}$$

$$n = 88 \quad \boldsymbol{R^2 = 0.1601}$$

3rd STEP: Using $R^2 = 0.1601$ compute test statistics: $F = 5.34$ with p-value=0.002, and $LM = 88 \times 0.1601 = 14.09$ (Chi-square with 3 dof, χ^2_3) with p-value=0.0028. **Strong evidence against homoscedasticity**.

Example: House Prices, hprice1.dta

As we mentioned earlier, using logarithmic transformation may reduce heteroscedasticity. Let us estimate a log-log specification (except rooms) for the house prices:

$$\widehat{\text{logprice}} = -\underset{(0.651)}{1.30} + \underset{(0.038)}{0.17}\ \text{loglotsize} + \underset{(0.093)}{0.70}\ \text{logsqrft} + \underset{(0.028)}{0.04}\ \text{bdrms}$$
$$n = 88 \quad R^2 = 0.643$$

Test statistics and associated p-values are

$$F = 1.141, \quad p - value = 0.245, \qquad LM = 4.22, \quad pvalue = 0.239$$

Obviously p-values are not small enough to reject the null. Therefore, we fail to reject the null hypothesis of homoscedasticity in the model with the logarithmic functional forms. In practice, using log transformation may ease the problem of heteroscedasticity.

White Test for Heteroscedasticity

- In White Test we assume that the constant variance assumption can be replaced with a weaker assumption: the squared error, "u^2, is uncorrelated with all the independent variables, x_j, their squares, x_j^2, and all the cross products, $x_j x_h$, $j \neq h$".
- This weaker assumption forms the basis of the White (1980) test for heteroscedasticity.
- The steps of this test is similar to the Breusch-Pagan test. The only difference is that squares and cross products of x variables are added to the test regression in the second step.

White Test for Heteroscedasticity

- For example, the test regression for $k = 3$ variables will be

$$\begin{aligned}\hat{u}^2 &= \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 \\ &+ \alpha_4 x_1^2 + \alpha_5 x_2^2 + \alpha_6 x_3^2 \\ &+ \alpha_7 x_1 x_2 + \alpha_8 x_1 x_3 + \alpha_9 x_2 x_3 + \nu\end{aligned}$$

- As k increases the degrees of freedom decreases significantly.
- Compared to the test regression of the Breusch-Pagan test, the White test regression contains 6 more parameters.
- The White test for heteroscedasticity uses the LM test statistic to test the following null hypothesis:

$$H_0: \ \alpha_1 = \alpha_2 = \ldots = \alpha_9 = 0$$

White Test for Heteroscedasticity

- This hypothesis can also tested using F test statistic. Both are asymptotically valid.
- When $k = 6$, the White test has 27 restrictions, using many degrees of freedom. This is a weakness of the White test.
- In practice, usually a simpler version of the White test, given below, is used. This does not lead to reduction in degrees of freedom as k increases.
- Instead of including squares and cross products directly in the test regression, one can use fitted values, $\hat{y}$, and their squares:

$$\hat{u}^2 = \alpha_0 + \alpha_1 \hat{y} + \alpha_2 \hat{y}^2 + \nu$$

- The null hypothesis of homoscedasticity becomes

$$H_0: \ \alpha_1 = \alpha_2 = 0$$

White Test for Heteroscedasticity

- The null hypothesis of homoscedasticity :

$$H_0: \ \alpha_1 = \alpha_2 = 0$$

- This can be tested using either F or LM test.
- No matter what k is, this test always has 2 restrictions, conserving on degrees of freedom.
- This test is especially useful when the variance is thought to change with the level of the expected value, $\mathrm{E}(y|x)$.

White Test: Example

- **1st STEP**: OLS estimation of model

$$\widehat{\text{lprice}} = -\underset{(0.651)}{1.30} + \underset{(0.038)}{0.17}\ \text{llotsize} + \underset{(0.093)}{0.70}\ \text{lsqrft} + \underset{(0.028)}{0.04}\ \text{bdrms}$$

$$n = 88 \quad R^2 = 0.643$$

- **2nd STEP**: Regression of $\hat{u}^2$ on $\hat{y}$ and $\hat{y}^2$

$$\widehat{\text{u}}^2 = \underset{(3.345)}{5.047} - \underset{(1.163)}{1.709}\,\widehat{lprice} + \underset{(0.100)}{0.145}\,\widehat{lprice}^2$$

$$n = 88 \quad R^2 = 0.03917$$

- **3rd STEP**: Calculate the test statistic: $LM = nR^2_{\hat{u}} = 88 \times 0.03917 = 3.447$, p-value=0.18. Decision: We fail to reject the null hypothesis of homoscedasticity.

Test for Heteroscedasticity

- Note that we assumed that the other classical assumptions are still valid (exogeneity, in particular).
- If these assumptions are not satisfied, for example, if the functional form is incorrect, then the tests for heteroscedasticity may reject the null hypothesis even if the variance is constant.
- In other words, the probability of Type 1. Error can be higher than the nominal significance level.
- This has led some economists to view tests for heteroscedasticity as general misspecification tests.
- But, there are better and more direct tests for functional form misspecification (eg Ramsey's RESET). We should first use these tests to rule out functional form misspecification since this is a more serious problem than heteroscedasticity.

Feasible GLS

- If $\boldsymbol{\Omega}$ is unconstrained there are $(n^2+n)/2$ additional parameters to estimate using just n observations. Obviously, this is impossible.
- FGLS procedures impose some structure on the model.
- For example, assume that the error variance is proportional in the following way:

$$\text{Var}(u_i|\boldsymbol{X}) = \sigma^2 \exp(\alpha_1 + \alpha_2 x_1 + \ldots + \alpha_k x_k) = \sigma^2 \exp(\boldsymbol{\alpha}^\top \boldsymbol{x}_i), \ \ i = 1, 2, \ldots,$$

- In this specification there are additional k parameters to estimate in the Ω matrix. This can easily be done.

Steps in Feasible Generalized Least Squares (FGLS)

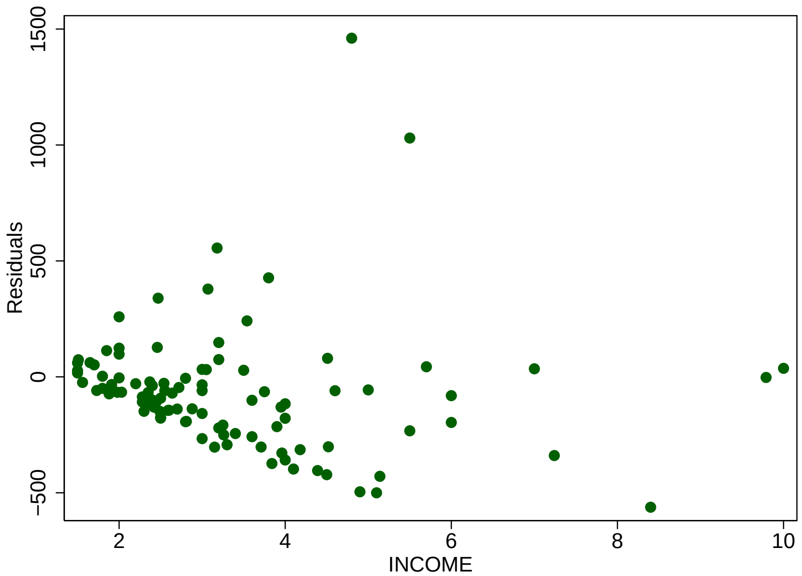
1. Estimate the model using OLS and save the residuals, $\hat{u}$
2. Compute the logarithm of the squared residuals $\Rightarrow \log(\hat{u}^2)$
3. Regress $\log(\hat{u}^2)$ on $x_1, x_2, \ldots, x_k$. Compute and save fitted values: $\hat{g}$
4. Exponentiate $\hat{g} \Rightarrow \hat{h} = \exp(\hat{g})$
5. Use $1/\hat{h}$ as weights and apply Weighted LS.

Note that we can also use NLS instead of log-linearizing the model.

Example: Credit Card Expenditures (Greene, p.269)

- Regression of credit card expenditures on a constant, age, income, income-squared and a dummy variable on whether an individual owns a house.
- Original data set includes 13144 individuals.
- Use a subset of this data with 72 observations (see Table F9-1)
- The following figure displays scatter of residuals and income.

Example: Credit Card Expenditures (Greene, p.269)



Estimating Covariance Matrix

- Recall that the asymptotic covariance matrix of the OLS estimator under nonspherical errors can be written as

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \frac{\sigma^2}{n}\left(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}X^\top\boldsymbol{\Omega}\boldsymbol{X}\right)\left(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}$$

- White heteroskedasticity consistent (robust) covariance matrix estimator is defined as follows:

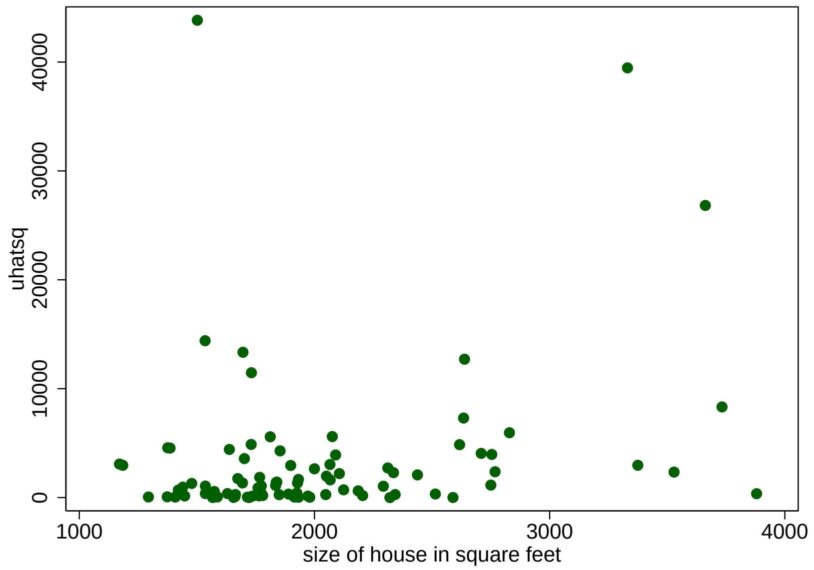
$$\text{Var}(\hat{\boldsymbol{\beta}}) = \frac{1}{n}\left(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2\boldsymbol{x}_i\boldsymbol{x}_i^\top\right)\left(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}$$

- This result implies that we can still use OLS framework, obtain the elements of the covariance matrix using the White (1980) formula. Standard inference procedures will be valid.
- Note that we do not need to specify exact form of the heteroskedasticity.
- Also note that this is a large sample estimator. May not be very useful in small samples.

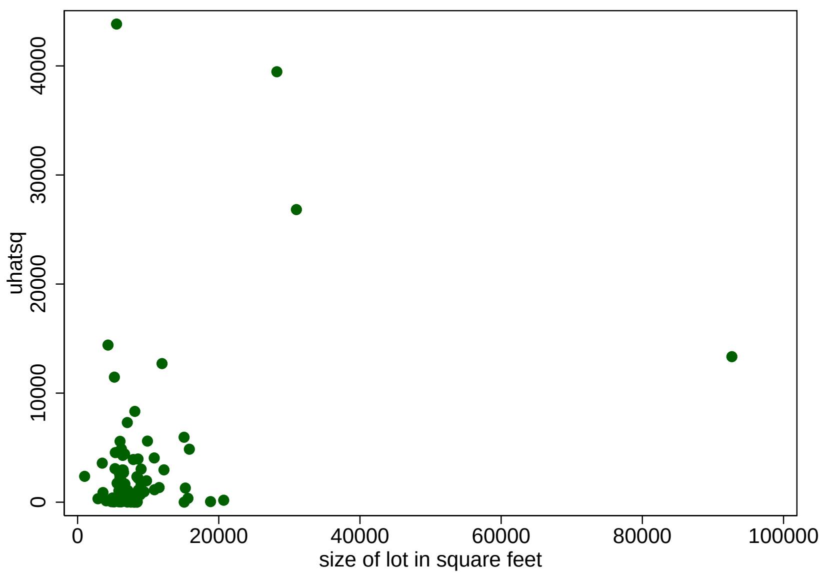
FGLS Application in STATA: hprice1.dta (Wooldridge)

```
reg price lotsize sqrft bdrms
predict uhat, resid
gen double uhatsq = uhat^2
scatter uhatsq sqrft
scatter uhatsq lotsize

. reg price lotsize sqrft bdrms

      Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  3,    84) =   57.46
       Model |  617130.701     3  205710.234           Prob > F      =  0.0000
    Residual |  300723.805    84   3580.0453           R-squared     =  0.6724
-------------+------------------------------           Adj R-squared =  0.6607
       Total |  917854.506    87  10550.0518           Root MSE      =  59.833

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lotsize |   .0020677   .0006421     3.22   0.002      .0007908    .0033446
       sqrft |   .1227782   .0132374     9.28   0.000      .0964541    .1491022
       bdrms |   13.85252   9.010145     1.54   0.128     -4.065141    31.77018
       _cons |  -21.77031   29.47504    -0.74   0.462     -80.38466    36.84405
------------------------------------------------------------------------------
```

House Prices: Squared residuals



House Prices: Squared residuals



FGLS Application in STATA: hprice1.dta (Wooldridge)

Heteroskedasticity tests:

```
. estat hettest  lotsize sqrft bdrms, iid

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: lotsize sqrft bdrms

         chi2(3)      =    14.09
         Prob > chi2  =   0.0028

. imtest, white

White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(9)      =     33.73
         Prob > chi2  =    0.0001

Cameron & Trivedi's decomposition of IM-test

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |      33.73      9    0.0001
            Skewness |       8.14      3    0.0432
            Kurtosis |       2.91      1    0.0882
---------------------+-----------------------------
               Total |      44.78     13    0.0000
---------------------------------------------------
```

FGLS Application in STATA: hprice1.dta (Wooldridge)

```
// NLS of uhatsq on exp(z'a)

. nl (uhatsq = exp({xb: lotsize sqrft bdrms one})), nolog
(obs = 88)

      Source |       SS       df       MS
-------------+------------------------------         Number of obs =        88
       Model |  1.9446e+09     4   486149632         R-squared     =    0.3597
    Residual |  3.4618e+09    84  41211951.3         Adj R-squared =    0.3292
-------------+------------------------------         Root MSE      =  6419.654
       Total |  5.4064e+09    88  61436391.4         Res. dev.     =  1788.652

------------------------------------------------------------------------------
      uhatsq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 /xb_lotsize |   .0000271   5.83e-06     4.64   0.000     .0000155    .0000387
   /xb_sqrft |   .0008292   .0002493     3.33   0.001     .0003334    .0013249
   /xb_bdrms |   .0941116   .1915099     0.49   0.624    -.2867268    .4749501
     /xb_one |   5.532212   .8229861     6.72   0.000     3.895614     7.16881
------------------------------------------------------------------------------
```


FGLS Application in STATA: hprice1.dta (Wooldridge)

```
. predict double varu, yhat          //sigmahat^2

. regress price lotsize sqrft bdrms [aweight=1/varu]  // FGLS
(sum of wgt is   4.2541e-02)

      Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  3,    84) =   22.75
       Model |  181931.623     3  60643.8745           Prob > F      =  0.0000
    Residual |  223910.762    84  2665.60431           R-squared     =  0.4483
-------------+------------------------------           Adj R-squared =  0.4286
       Total |  405842.385    87    4664.855           Root MSE      =  51.629

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lotsize |   .0034323   .0012815     2.68   0.009     .0008839    .0059807
       sqrft |   .0955063   .0156009     6.12   0.000     .0644821    .1265304
       bdrms |   8.137567   8.361846     0.97   0.333    -8.490881    24.76602
       _cons |   38.94434   31.37514     1.24   0.218    -23.44857    101.3372
------------------------------------------------------------------------------
```