## 8 Heteroskedasticity

What does ie mean to have homoskedastic errors:

$$Var(u_i|x_i) = \sigma$$

What does it mean to have heteroskaedastic errors:

$$Var(u_i|x_i) = f(\sigma_i, x_i)$$

i.e. the conditional variance of the error term depends on the independent variables used in regression.

## 8.1 Consequence of Heteroskedasticity for OLS

1. Note that under MLR.1 to MLR.4 OLS estimates are unbiased and consistent. It has nothing to do with homoskedasticity assumption.

   Why: because both variances in population R-squared are unconditional variances, the population R-squared is unaffected by the presence of heteroskedasticity.

2. Where is the problem: estimator of variances of $\hat{\beta}$ are biased which makes the standard errors for confidence intervals incorrect.

   + Issues in inference of F or the LM statistic under heteroskedasticity.

## 8.2 Heteroskedasticity-Robust Inference after OLS Estimation

1. Problem: how do we adjust t, F and LM statistics to make the correct inference? This is under any form of Heteroskedasticity.

2. *Solution*: compute standard errors which take into account the possibility of heteroskedasticity.

3. Implementation in a single linear model:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Heteroskedasticity implies that $u_i$ varies over $i$ conditional on the $x_i$. This implies the following in the simplest of cases.

$$Var(u_i|x_i) = \sigma_i^2$$

4. OLS estimator (Recall 2.52 page 43, JW)

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \beta_1 + \frac{\sum_{i=0}^{n}(x_i - \bar{x})u_i}{SST_x}$$

5. Only using MLR.1 to MLR.4 we know the variance of the parameter estimate to be

$$Var(\hat{\beta}_1) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2\sigma_i^2}{SST_x^2}$$

When homoskedasticity implies $\sigma_i^2 = \sigma^2$, we can take the $\sigma$ out of summation to simplify as:

$$Var(\hat{\beta}_1) = \sigma^2 \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{SST_x^2} = \sigma^2 \frac{1}{SST_x}$$

6. Under heteroskedasticity, White showed that an estimate of Var($\hat{\beta}$) is given by the following:

$$Var(\hat{\beta}) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \hat{u}_i^2}{SST_x^2}$$

where $\hat{u}_i$ is the OLS residual from initial regression of $y$ on $x$.

White showed that

$$plim_{n\to\infty} \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \hat{u}_i^2}{SST_x^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sigma_i^2}{SST_x^2}$$

7. *In Multiple Regression Model*: this can be implemented by obtaining the $\hat{u}_i^2$ from the following regression:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

The square root of the unconditional variance is called the Heteroskedasticity-robust standard error:

$$\widehat{Var}(\hat{\beta}_j) = \frac{\sum_{i=1}^{n} \hat{r}_{ij}^2 \hat{u}_i^2}{SSR_j^2}$$

where $\hat{r}_{ij}$ denoted the $i^{th}$ residual from regressing $x_j$ on all other independent variables and $SSR_j$ is the sum of squared residuals from this regression.

8. There are many adjustments to the standard error and method of computing it, but all are equivalent asymptotically or in large samples.

   We can use this to come up with correct $t$-statistics, F, LM and hence correct inference in the presence of heteroskedasticity

   - ▷ Heteroskedasticity Robust $t$-Statistic is given as:

$$t = \frac{\text{estimate - hypothesized value}}{\text{standard error}}$$

9. Question: If Heteroskedasticity-robust standard errors (HRSE) are always correct, should we always use them?

   No, for small samples normal SE are better. Why? Because HRSE corrects for heteroskedasticity. If there is no hetero in the data in the first place, we are correcting for something which might induce an error in our inference.

10. Heteroskedasticity-robust F-statistic: is also called the Wald statistic can be obtained for *any* form of heteroskedasticity $\Longrightarrow$ derivation is involved and we wont need it for the moment.

    - ▷ Similarly, remember Chow test of common coefficient across two groups? [Section 7.4c: TO DO computer exercise C14]

### 8.2.1 Computing Heteroskedasticity Robust LM Tests

1. To compute LM consistent with hetero, consider the following model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + u$$

2. We want to test whether $\beta_4 = 0, \beta_5 = 0$.

3. Estimate the restricted model and save the residuals $\tilde{u}$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \tilde{u}$$

4. For each excluded variable $x_{ei}$, regress them on the remaining included variable $x_{ii}$ and save the residuals as $r_{ei}$. This under (2) is as follows:

$$x_4 \text{ on } x_1, x_2 \text{ and } x_3 \text{ to obtain } r_4$$
$$x_4 \text{ on } x_1, x_2 \text{ and } x_3 \text{ to obtain } r_5$$

5. Regress the following and obtain SSR from the regression.

$$1 \text{ on } r_4\tilde{u} \text{ and } r_5\tilde{u} \text{ to obtain SSR}$$

6. LM statistic is $n - SSR$

## 8.3 Testing for Heteroskedasticity

1. Do we really need to know whether there is heteroskedasticity to implement the methods? *No*, but it helps to have a method.

2. Idea: We want to test whether the regression error, conditional on the $x$'s is homoskedastic or heteroskedastic. Let's start with a model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

Under MLR.1 to MLR.4 we have consistent and unbiased estimates. We add MLR.5

$$H_0 : Var(u|x_1, x_2, x_3, \ldots) = \sigma^2$$

3. If $H_0$ is false, then the conditional expectation function of $u$ can be any function of any of the $x_k$. Let us assume it is a linear function of the following form:

$$u^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k + v$$

4. Null hypothesis under homoskedasticity implies:

$$H_0 : \delta_0 = \delta_1 = \ldots = \delta_k = 0$$

5. But we dont know the true $u^2$ so we use the sample proxy to obtain:

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k + v$$

6. The F statistic is:

$$F = \frac{R^2_{\hat{u}^2}/k}{(1 - R^2_{\hat{u}^2})/(n - k - 1)}$$

where k are the number of regressors in step 5. F stat has approximately $F_{k,n-k-1}$ distribution under the null hypothesis.

7. LM statistic is called the **Breusch-Pagan test** for heteroskedasticity:

$$LM = n \cdot R^2_{\hat{u}^2}$$

8. **Form of heteroskedasticity**: If we think that variance of the error term depends upon only some of the $x^s$ , we repeat step 5 but only on selected $x$ and construct the statistic.

### 8.3.1 White Test for Heteroskedasticity

1. Idea: White suggested to test different form of heteroskedasticity by including squared terms and cross products in the following equation:

$$\begin{aligned}\hat{u}^2 =& \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 + \delta_4 x_1^2 + \delta_5 x_2^2 + \delta_6 x_3^2 \\ &+ \delta_7 x_1 x_2 + \delta_8 x_1 x_3 + \delta_9 x_2 x_3 + v\end{aligned}$$

2. The null hypothesis for heteroskedasticity implies that all $\delta$'s are zero except $\delta_0$, thus 9 exclusion restrictions.

   Naturally, it uses a lot of degrees of freedom even with a small size of the model.

We could alternatively use OLS fitted values to get around the degree of freedom issue by running the following model:

$$\begin{aligned}\hat{u}^2 =& \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + v \\ \hat{y} =& \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3\end{aligned}$$

In order to test heteroskedasticity, we merely test whether $\delta_0$ and $\delta_1$ are zeros or not and conserve degrees of freedom.

3. **Caution**: If MLR.4 is violated - in particular if the functional form of $E(y|\mathbf{x})$ is misspecified - then a test for heteroskedasticity can reject the null $H_0$ even if $Var(y|\mathbf{x})$ is homoskedastic.

## 8.4 Weighted Least Square Estimation

1. **Idea**: If you could not certainly tell whether there is heteroskedasticity, then you could alternatively weight your OLS estimates according to the form of heteroskedasticity.

### 8.4.1 Heteroskedasticity is Known to a Multiplicative Constant

Consider the scenario when you know the functional form $h(\mathbf{x})$ of conditional variance with an unknown population variance $\sigma^2$

$$Var(u|\mathbf{x}) = \sigma^2 h(\mathbf{x})$$

For instance, you want to know the effect of individual savings on income, but you suspect that conditional error variance is linearly related to income of the individual.

$$sav_i = \beta_0 + \beta_i inc_i + u_i \qquad Var(u_i|inc_i) = \sigma^2 inc_i$$

Idea: We use our knowledge of the conditional variance and transform the original regression

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + u_i$$

into one with homoskedastic errors.

We do this by dividing $\frac{u_i}{\sqrt{h_i}}$ such that the conditional variance of this transformed error is

$$E\left[\frac{u_i}{\sqrt{h_i}}\right] = \frac{\sigma^2 h_i}{h_i} = \sigma^2$$

We obtain this transformed error from the transformed regression of the following form (remember chapter 6)

$$\frac{y_i}{\sqrt{h_i}} = \frac{\beta_0}{\sqrt{h_i}} + \frac{\beta_1}{\sqrt{h_i}} x_{i1} + \frac{\beta_2}{\sqrt{h_i}} x_{i2} + \ldots + \frac{\beta_k}{\sqrt{h_i}} x_{ik} + u_i$$

In our savings example:

$$\frac{sav_i}{\sqrt{inc_i}} = \frac{\beta_0}{\sqrt{inc_i}} + \frac{\beta_1 inc_i}{\sqrt{inc_i}} + \frac{u_i}{\sqrt{inc_i}}$$

This is an example of the **Generalized Least Square estimation of a model** whose original version had heteroskedasticity. The transformed equation satisfies all the Gauss-Markov assumptions

1. **MLR.1** : linear in parameters. The estimated parameters in this case are $\beta_0^* = \frac{\beta_0}{\sqrt{inc_i}}$. The model is linear in the transformed parameters.

2. **MLR.2**: Random sampling merely implies that when we are recording income and savings of people, they are representative of the population.

3. **MLR.3**: No perfect collinearity which in this case comes automatically cause there is no other variable.

4. **MLR.4**: Zero conditional mean implies that the transformed error from the regression has zero mean. Since its from the regression, its already conditioned on income.

5. **MLR.5**: Homoskedasticity i.e. $Var(u_i|\mathbf{x}_i) = \sigma^2$

6. **MLR.6**: Normality i.e. $u_i \ N(0, \sigma^2)$

Does it help inference? Yes since the transformed estimates follow all the MLR assumptions. But interpret the original estimates. This particular GLS is the *weighted least square* since the weights in this case are $\frac{1}{h_i}$.

**Idea**: WLS gives less weight to observations with higher variance. A weighted least squares estimator can be defined for any set of positive weights. OLS is the special case that gives equal weight to all observations.

*Best procedure*: is when you weight each squared residual by the inverse of the conditional variance of $u_i$ given $\mathbf{x}_i$

## Example 8.6: Financial Wealth Equation

Consider an ordinary least square estimation for the following equation:

$$nettfa = \beta_0 + \beta_1 inc + u$$

```
> summary(lm(nettfa ~ inc, k401ksubs))

Call:
lm(formula = nettfa ~ inc, data = k401ksubs)

Residuals:
    Min      1Q  Median      3Q     Max
-504.39  -18.10   -4.29    6.73 1475.04

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -20.17948    1.17643  -17.15   <2e-16 ***
inc           0.99991    0.02554   39.15   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 59.26 on 9273 degrees of freedom
Multiple R-squared:  0.1418,    Adjusted R-squared:  0.1417
F-statistic:  1532 on 1 and 9273 DF,  p-value: < 2.2e-16
```

If we use data on only single people to know their motivations for savings, we do the following:

```
> mod = lm(nettfa ~ inc, data = k401ksubs, subset = (fsize==1))
> summary(mod)

Call:
lm(formula = nettfa ~ inc, data = k401ksubs, subset = (fsize ==
    1))

Residuals:
    Min      1Q  Median      3Q     Max
-185.12  -12.85   -4.85    1.78 1112.66

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -10.5709     2.0607   -5.13 3.18e-07 ***
inc           0.8207     0.0609   13.48  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 45.59 on 2015 degrees of freedom
Multiple R-squared:  0.08267,   Adjusted R-squared:  0.08222
F-statistic: 181.6 on 1 and 2015 DF,  p-value: < 2.2e-16
```

Now we can report heteroskedasticity consistent standard errors instead of the usual standard errors if we suspect the possibility of heteroskedasticity.

```
> coeftest(mod, vcov = vcovHC(mod, type="HC1"))

t test of coefficients:

             Estimate Std. Error t value  Pr(>|t|)
(Intercept) -10.57095    2.53027 -4.1778 3.069e-05 ***
inc           0.82068    0.10359  7.9221 3.826e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Alternatively we can run a weighted least squares regression if we

```
> mod2 = lm(nettfa ~ inc, data = k401ksubs, subset = (fsize==1), weights = 1/inc)
> summary(mod2)

Call:
lm(formula = nettfa ~ inc, data = k401ksubs, subset = (fsize ==
    1), weights = 1/inc)

Weighted Residuals:
    Min      1Q  Median      3Q     Max
-23.469  -2.339  -1.086   0.352 178.220

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -9.58070    1.65328  -5.795 7.91e-09 ***
inc          0.78705    0.06348  12.398  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.219 on 2015 degrees of freedom
Multiple R-squared:  0.07088,   Adjusted R-squared:  0.07042
F-statistic: 153.7 on 1 and 2015 DF,  p-value: < 2.2e-16
```

We suspect that savings to 401K start only after individuals reach a certain age after which such contributions increase. We suspect a quartratic relation for all age range which increases with an upward slope after age 25.

```
> mod3 = lm(nettfa ~ inc + I((age - 25) * (age-25)) + male + e401k, data = k4
01ksubs, subset = (fsize==1))
> summary(mod3)

Call:
lm(formula = nettfa ~ inc + I((age - 25) * (age - 25)) + male +
    e401k, data = k401ksubs, subset = (fsize == 1))

Residuals:
    Min      1Q  Median      3Q     Max
-176.04  -14.17   -3.15    6.01 1111.42

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)               -20.984990   2.472022  -8.489   <2e-16 ***
inc                         0.770583   0.061452  12.540   <2e-16 ***
I((age - 25) * (age - 25))  0.025127   0.002593   9.689   <2e-16 ***
male                        2.477927   2.047776   1.210   0.2264
e401k                       6.886223   2.123275   3.243   0.0012 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 44.49 on 2012 degrees of freedom
Multiple R-squared:  0.1279,    Adjusted R-squared:  0.1261
F-statistic: 73.75 on 4 and 2012 DF,  p-value: < 2.2e-16
```

A WLS version of the expanded model is

```
> mod4 = lm(nettfa ~ inc + I((age - 25) * (age-25)) + male + e401k, data = k4
01ksubs, subset = (fsize==1), weights = 1/inc)
> summary(mod4)

Call:
lm(formula = nettfa ~ inc + I((age - 25) * (age - 25)) + male +
    e401k, data = k401ksubs, subset = (fsize == 1), weights = 1/inc)

Weighted Residuals:
    Min      1Q  Median      3Q     Max
-26.613  -2.491  -0.803   0.934 178.052

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                -16.702521   1.957995  -8.530  < 2e-16 ***
inc                          0.740384   0.064303  11.514  < 2e-16 ***
I((age - 25) * (age - 25))   0.017537   0.001931   9.080  < 2e-16 ***
male                         1.840529   1.563587   1.177  0.23929
e401k                        5.188281   1.703426   3.046  0.00235 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.065 on 2012 degrees of freedom
Multiple R-squared:  0.1115,    Adjusted R-squared:  0.1097
F-statistic: 63.13 on 4 and 2012 DF,  p-value: < 2.2e-16
```

Two cases when WLS works better than OLS:

1. Consider a model of employee ($e$) contribution to a savings fund in a firm ($i$). In this case, even with homoskedastic errors if you dont have employee level data for all firms, then you could work with average contribution of employees in one firm to examine cross firm behavior. But running an OLS would be wrong cause the error in the second model are not homoskedastic:

$$contrib_{ie} = \beta_0 + \beta_1 earns_{ie} + \beta_2 age_{ie} + \beta_3 mrate_{ie} + u$$
$$\bar{contrib}_{ie} = \beta_0 + \beta_1 \bar{earns}_{ie} + \beta_2 \bar{age}_{ie} + \beta_3 \bar{mrate}_{ie} + u$$

2. Similar issue arises when using per capita data at the city, county, state or country level. If the individual-level equation satisfies the Gauss-Markov assumptions, then the error in the per capita equation has a variance proportional to one over the size of the population. Therefore, weighted least squares with weights equal to the population is appropriate.

If heteroskedasticity exists at the individual level, then the proper weighting depends on the form of heteroskedasticity

### 8.4.2 Unknown Heteroskedasticity: Feasible GLS

What if we do not know the exact form of the heteroskedasticity? Instead of using heteroskedasticity robust standard errors, one could use a version of the GLS estimator called the Feasible GLS.

Estimate conditional variance of the error term from data:

$$Var(u|\mathbf{x}) = \sigma^2 exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k)$$

We could have used a linear model of conditional variance but they cause predicted variance to be negative.

How do we estimate this model?

$$u^2 = \sigma^2 exp(\delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k)v$$

We transform this model by taking log to estimate the following model:

$$log(u^2) = \alpha_0 + \delta_1 x_1 + \cdots + \delta_k x_k + e$$

We can now use the fitted values from the above regression $\hat{h}_i = exp(\hat{g}_i)$ to obtain weights $\frac{1}{\hat{h}_i}$ on the WLS regression.

Having to estimate $h_i$ using the same data means that the FGLS estimator is no longer unbiased (so it cannot be BLUE, either). Nevertheless, the FGLS estimator is consistent and asymptotically more efficient than OLS. This means that if you have a large sample, the FGLS estimator would do a better job than OLS in precision of estimates.

F-stat after WLS: The only caution we must take is that the weights should be identical in both the restricted and unrestricted regression.

Difference between OLS and FGLS: they tend to give different estimates, but we would not be worried unless the results are overturned. Its best to run both models to get a sense of the magnitude and direction of the estimates.

- If they differ, then our model is not clearly identified i.e. the other gauss markov assumptions are not met. Always check whether $E(y|\mathbf{x}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$ i.e. the MLR.4 is satisfied or not.
- There could be functional form mis-specification too.

### 8.4.3 What if we assume wrong Heteroskedasticity Function

The key issue here is whether the misspecification of $h(x)$ causes bias or inconsistency in the WLS estimator.

Answer: It doesn't if MLR.4 assumption is valid.

The logic is that if $E(u|\mathbf{x}) = 0$, then any functional transformation of the error is also zero and weighting is a functional transformation.

Implications of incorrect heteroskedasticity function:

1. Standard errors are incorrect even for large samples. *Solution*: obtain Heteroskedasticity consistent standard errors for the WLS too!

2. WLS is inefficient in comparison to OLS. *However, practically* - it is often better to use a wrong form of heteroskedasticity and apply WLS than to ignore heteroskedasticity altogether in estimation and use OLS.

### 8.4.4 Prediction Intervals with Heteroskedasticity

1. Prediction intervals get affected by heteroskedasticity but not the predictions themselves. In order to obtain the prediction, we extend the methods explored earlier by estimating the following regression under WLS

$$y_i = \theta_0 + \beta_1(x_{i1} - x_1^0) + \cdots + \beta_k(x_{ik} - x_k^0) + u_i$$

   From this we get $\hat{y}^0 = \hat{\theta}_0$ which is the in-sample prediction with its standard error as $se(\hat{\theta}_0)$

2. For out of sample prediction, we need $sd(u^0)$. Under our assumption of variance $Var(u^0|\mathbf{x} = \mathbf{x}^0) = \sigma^2 h(\mathbf{x}_0)$ we can obtain a standard error by replacing the population variance with its estimated sample counterpart.

3. *FGLS and prediction intervals*: IF we have to estimate variance function as in FGLS, then we cannot obtain exact intervals. The next best option is to use $\hat{h}(\mathbf{x}_0)$

4. JW proceeds with explaining how to obtain prediction intervals for a log model of the following form:

$$log(y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

where $u|\mathbf{x} \sim N[0, exp(\delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k)]$

(a) The CEF (Conditional expectation function) is

$$E(y|\mathbf{x}) = exp(\beta_0 + \mathbf{x}\boldsymbol{\beta} + exp(\delta_0 + \mathbf{x}\boldsymbol{\delta})/2)$$

which is estimated using WLS with estimated variance as in FGLS.

(b) The estimated variance under FGLS is $\hat{\sigma}^2 exp(\hat{g}_i) = \hat{\sigma}^2 exp(\hat{h}_i)$ and the fitted value is

$$\hat{y}_i = exp(\widehat{log(y)}_i + \hat{\sigma}^2 exp(\hat{g}_i))$$

Prediction intervals can be obtained by using the strategy adopted previously for level models with adjustments for log dependent variables.

## 8.5 Linear Probability Model: Revisited

1. **LPM model with heteroskedasticity**: simplest solution is to compute heteroskedasticity robust standard errors. Recall: this is when we dont know the form of heteroskedasticity.

   (a) Recall the response probability:

   $$p(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

   (b) For each observsation $i$, $Var(y_i|x_i)$ is estimated by OLS fitted values as follows:

   $$Var(y_i|x_i) = \hat{y}_i(1 - \hat{y}_i) = \hat{h}_i$$

   However, when the fitted values are negative or greater than 1 the variance would be negative which does not help. In such instances, just abandon WLS and use heteroskedasticity robust standard error.