Multiple Regression Analysis: Heteroskedasticity. Wooldridge (2013), Chapter 8.

- What is Heteroskedasticity? Why Worry About Heteroskedasticity?
- Variance of the OLS estimator with Heteroskedasticity
- Robust Standard Errors
- Heteroskedastic-robust Wald statistic and A Robust Lagrange Multiplier Statistic
- Testing for Heteroskedasticity (The Breusch-Pagan Test, The White Test)
- Weighted Least Squares, Generalized Least Squares, Feasible GLS
- Prediction and Prediction Intervals with Heteroskedasticity

Recall the *Gauss-Markov* assumptions:

1. Population model is linear in parameters:
   $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$.
2. We can use a random sample of size n,$\{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i) : i = 1, 2, \ldots, n\}$, from the population model, so that the sample model is

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

3. $E(u|x_1, x_2, \ldots\ x_k) = 0$.
4. None of the x's is constant, and there are no exact linear relationships among them (no perfect *multicollinearity*).
5. *Homoskedasticity* implied that conditional on the explanatory variables, the variance of the unobserved error, u, was constant $Var(u|x_1, ..., x_k) = \sigma^2$.

- If 5 is not true, that is if the conditional variance of u is different for different values of the x's, then the errors are *heteroskedastic*.
- Notice that
$$Var(u|x_1, ..., x_k) = Var(y|x_1, ..., x_k).$$
- Is the assumption of *homoskedasticity* realistic?
- In cross-sectional data usually the errors are *heteroskedastic*, that is $Var(u|x_1, ..., x_k)$ varies with the regressors.

Example:
Let us consider the regression model

$$\log(wage) = \beta_0 + \beta_1 female + u,$$

where female is a dummy variable:

$$female = \begin{cases} 1 & \text{if female} \\ 0 & \text{otherwise} \end{cases}.$$

Homoskedasticy implies that

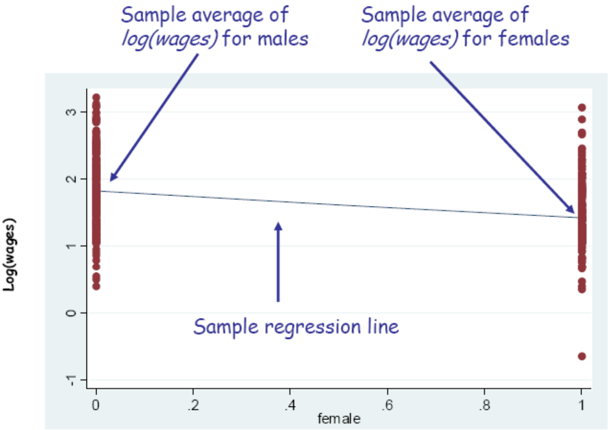
$$Var(\log(wage)|female = 1) = Var(\log(wage)|female = 0)$$

which is equivalent to

$$Var(\log(wage)|females) = Var(\log(wage)|males).$$

Example: Let us consider the a sample taken from the 1976 US Current Population Survey ($n = 526$).

ScatterPlot

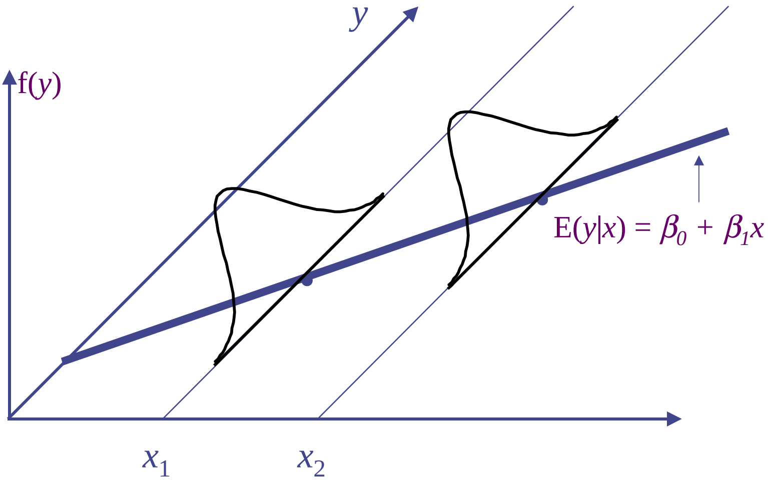


Example: Sample regression line:

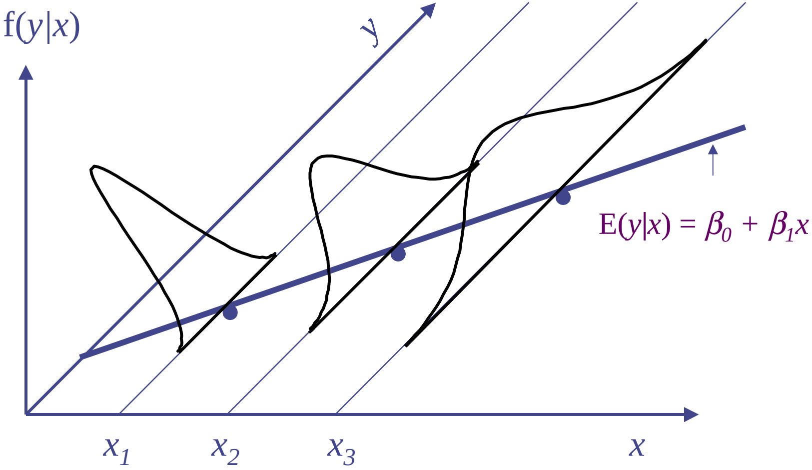
$$\widehat{\log(wage)} = 1.8136 - 0.3972 female$$

- The sample variance of $\log(wage)$ for males is 0.28602.
- The sample variance of $\log(wage)$ for females is 0.19734.
- This is an indication that the assumption of Homoskedasticity does not hold.

- *Homoskedastic Case*: $Var(u|x) = \sigma^2$.



- *Heteroskedastic Case*: $Var(u|x)$ varies with x.



- OLS is still *unbiased* and *consistent*, even if we do not assume homoskedasticity.
- Now OLS is *not BLUE*.
- The *standard errors* of the estimates proposed before are *biased* if we have heteroskedasticity and hence not valid.
- If the standard errors are biased, we *cannot use* the usual t statistics or F statistics or LM statistics for drawing inferences.
- We have to propose standard errors that are valid even under heteroskedasticity.

Consider the simple linear regression model

$$y = \beta_0 + \beta_1 x + u,$$
$$E(u|x) = 0, Var(u|x) = \sigma^2(x).$$

For this model

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

Hence conditional on $x_1, ..., x_n$

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

$SST_x = \sum_{i=1}^{n}(x_i - \bar{x})^2$. Notice that $\sigma^2(x_i)$ is unknown.
A valid estimator for $Var(\hat{\beta}_1)$ is

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

where $\hat{u}_i$ are the OLS residuals.
Remark: $E(u^2|x) = \sigma^2(x)$.

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + u,$$
$$E(u|\mathbf{x}) = 0, Var(u|\mathbf{x}) = \sigma^2(\mathbf{x}),$$
$$\mathbf{x} = (x_1, ..., x_k).$$

For the general multiple regression model, a valid estimator of $Var(\hat{\beta}_j)$ with heteroskedasticity is

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

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

- Now that we have a consistent estimate of the variance, the square root can be used as a standard error for inference, that is $se(\hat{\beta}_j) = \sqrt{\widehat{Var}(\hat{\beta}_j)}$.
- Typically call these *robust standard errors* or *White, Huber or Eicker standard errors*.
- Once the heteroskedastic robust standard errors are obtained the heteroskedastic-robust t statistic is computed in the usual way

$$t = \frac{estimator - hypothesized\ value}{standard\ error}.$$

- One can show that $t \overset{a}{\sim} N(0,1)$.
- These robust standard errors only have asymptotic justification – with small sample sizes t statistics formed with robust standard errors will *not* have a distribution close to the t, and inferences will not be correct.

Consider the multiple regression model

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + u.$$

Suppose that we would like to test $H_0 : \beta_1 = \beta_2 = ... = \beta_q = 0$.

- It is possible to obtain F and LM statistics that are robust to heteroskedasticity of an unknown arbitrary form.
- The heteroskedastic robust F statistic (or a simple transformation of it) is called a *heteroskedastic-robust Wald statistic*.
- The specific formula of the Wald statistic requires matrix algebra and will not be given here, though most of the Econometric software have procedures to compute it.
- Since we are testing the validity of q restrictions, the asymptotic distribution of the heteroskedastic-robust Wald statistic is $\chi^2(q)$.
- This can also used to test other types of restrictions on the parameters.

Regressing $\log(wage)$ on education experience and tenure we obtain (n=526):

Regressors	Estimates	Usual Std. Err.	Robust Std. Err.
Intercept	0.28436	0.10419	0.11171
education	0.09203	0.00733	0.00792
experience	0.00412	0.00172	0.00175
tenure	0.02207	0.00309	0.00378

Tests of joint zero restrictions on *exper* and *tenure*:

- Value of the usual F-Statistic $F^{act} = 49.6852$ $(F \sim F(2, 522))$
- Value of the heteroskedastic-robust Wald statistic: $W^{act} = 74.1037$. $(W \sim \chi^2(2))$

Suppose that we would like to test $H_0 : \beta_1 = \beta_2 = ... = \beta_q = 0$.

1. Run OLS on the restricted model and save the residuals $\hat{u}$.
2. Regress each of the excluded variables on all of the included variables (q different regressions) and save each set of residuals $\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_q$.
3. Regress a variable defined to be $= 1$ on $\hat{r}_1\hat{u}, \hat{r}_2\hat{u}, \ldots, \hat{r}_q\hat{u}$, with **no** intercept.
4. The LM statistic is $n - SSR_1$, where SSR_1 is the sum of squared residuals from this final regression.
5. Under H_0, $LM \overset{a}{\sim} \chi^2(q)$.

Let $\mathbf{x} = (x_1, x_2, ..., x_k)$ and consider the linear regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u,$$
$$E(u|\mathbf{x}) = 0.$$

- There is heteroskedasticity if $Var(u|\mathbf{x}) = \sigma^2(\mathbf{x})$.
- There is homoskedasticity if $Var(u|\mathbf{x}) = \sigma^2$.

- Essentially want to test $H_0 : Var(u|x_1, x_2, \ldots, x_k) = \sigma^2$, which is equivalent to $H_0 : E(u^2|x_1, x_2, \ldots, x_k) = E(u^2) = \sigma^2$.
- If assume the relationship between u^2 and x_j will be linear, can test as a linear restriction.
- So, for $u^2 = \delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k + v$, $E(v|x_1, x_2, \ldots\ x_k) = 0$, this means testing $H_0 : \delta_1 = \delta_2 = \ldots = \delta_k = 0$.

- Don't observe the error, but can estimate it with the residuals from the OLS regression, that is we replace u by $\hat{u}$.
- After regressing the residuals squared on all of the x's, can use the R^2 to form an F or LM test.
- The F statistic is just the reported F statistic for overall significance of the regression, $F = [R^2/k]/[(1-R^2)/(n-k-1)]$, which is distributed $F(k, n-k-1)$.
- The LM statistic is $LM = nR^2$, which is distributed $\chi^2(k)$.

Example 1: Consider the following regression, where *Res*2 are the squares of the residuals of the regression of $\log(wages)$ on *female*.Test the null hypothesis of homoskedasticity at 5% level.

$$\begin{aligned}\widehat{Res2} &= 0.2850 - 0.0884 female, \\ R^2 &= 0.012744, \\ n &= 526.\end{aligned}$$

Example 2: Let *Res*2 be the squares of the residuals of the regression of $\log(wages)$ on an intercept, *educ*, *exper* and *tenure* ($n = 526$). We run the regression of *Res*2 on an intercept, *educ*, *exper* and *tenure* and obtain $R^2 = 0.0205$.Test the null hypothesis of homoskedasticity at 5% level.

- The *Breusch-Pagan test* will detect any *linear* forms of heteroskedasticity.
- The *White test* allows for *nonlinearities* by using squares and crossproducts of all the x's, that is, we run the regression

$$\begin{aligned}\hat{u}^2 &= \delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k + \delta_{k+1} x_1^2 + \ldots + \delta_{2k} x_k^2 + \delta_{2k+1} x_1 x_2 + \\ &\quad ...\delta_{k+k(k+1)/2} x_k x_{k-1} + error\end{aligned}$$

Want to test $H_0 : \delta_1 = \delta_2 = \ldots = \delta_{k+k(k+1)/2} = 0$.

- The F statistic is just the reported F statistic for overall significance of the regression, $F = [R^2/q]/[(1-R^2)/(n-q-1)]$, which is distributed $F(q, n-q-1)$, where $q = k + k(k+1)/2$.
- The LM statistic is $LM = nR^2$, which is distributed $\chi^2(q)$.
- **Example:** if $k = 3$, we run the regression of $\hat{u}^2$ on an intercept and $x_1, x_2, x_3, x_1^2, x_2^2, x_3^2, x_1x_2, x_2x_3, x_1x_3$.
- We are testing if 9 parameters are equal to zero so $F \overset{a}{\sim} F(9, n-9-1)$ and $LM \overset{a}{\sim} \chi^2(9)$.

Example: Let $Res2$ be the squares of the residuals of the regression of $\log(wages)$ on an intercept, $educ$, $exper$ and $tenure$ ($n = 526$). We run the regression of $Res2$ on an intercept, $educ$, $exper$, $tenure$, $educ^2$, $exper^2$, $tenure^2$, $educ \times exper$, $educ \times tenure$ and $exper \times tenure$ and obtain $R^2 = 0.0394$. Test the null hypothesis of homoskedasticity at 5% level.

The power of the White test is usually not high because we are testing the significance of a large number of regressors. For instance if $k = 6$ we are testing the significance of 27 regressors.
A possible remedy is to drop the cross terms, if $k = 6$ we are testing the significance of 12 regressors.

- In this case we run the regression

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k + \delta_{k+1} x_1^2 + \ldots + \delta_{2k} x_k^2 + error$$

Want to test $H_0 : \delta_1 = \delta_2 = \ldots = \delta_{2k} = 0$.

- The F statistic is just the reported F statistic for overall significance of the regression, $F = [R^2/2k]/[(1 - R^2)/(n - 2k - 1)]$, which is distributed $F(2k, n - 2k - 1)$.
- The LM statistic is $LM = nR^2$, which is distributed $\chi^2(2k)$.

Example: Let $Res2$ be the squares of the residuals of the regression of $\log(wages)$ on an intercept, $educ$, $exper$ and $tenure$ ($n = 526$). We run the regression of $Res2$ on an intercept, $educ$, $exper$, $tenure$, $educ^2$, $exper^2$ and $tenure^2$ and obtain $R^2 = 0.0268$. Test the null hypothesis of homoskedasticity at 5% level.

An alternative is the following:

- Consider that the fitted values from OLS, $\hat{y}$, are a function of all the x's.
- Thus, $\hat{y}^2$ will be a function of the squares and crossproducts and $\hat{y}$ and $\hat{y}^2$ can proxy for all of the x_j, x_j^2, and $x_j x_h$.
- Regress the residuals squared on $\hat{y}$ and $\hat{y}^2$ and use the R^2 to form an F or LM statistic. In this case $F \overset{a}{\sim} F(2, n-2-1)$ and $LM \overset{a}{\sim} \chi^2(2)$.
- Note only testing for 2 restrictions now.

Example: Let $Res2$ and $\widehat{\log(wages)}$ be the squares of the residuals and the fitted values, respectively, of the regression of $\log(wages)$ on an intercept, $educ$, $exper$ and $tenure$ ($n = 526$). We run the regression of $Res2$ on an intercept, $\widehat{\log(wages)}$ and $\left(\widehat{\log(wages)}\right)^2$ and obtain $R^2 = 0.0127$. Test the null hypothesis of homoskedasticity at 5% level.

- Under heteroskedasticity the usual formula for the standard errors is not valid.
- We need to compute robust standard errors, that are consistent under heteroskedasticity.
- Once the heteroskedastic robust standard errors are obtained the heteroskedastic-robust t statistic is computed in the usual way

$$t = \frac{estimator - hypothesized\ value}{standard\ error}.$$

- The F and LM statistics introduced under the Gauss-Markov Assumptions are also not valid.
- The heteroskedastic robust F statistic (or a simple transformation of it) is called a *heteroskedastic-robust Wald statistic* .
- One can also use a LM test statistic which is valid under heteroskedasticity.

- We can test for Homoskedasticity by testing the joint significance of the independent variables in the regression of the squared residuals $\hat{u}_i^2$ on:
    - All the regressors.
    - All the regressors, squares of the regressors and crossproducts.
    - All the regressors, squares of the regressors.
    - The fitted values and squares of the fitted values.

- We can always estimate robust standard errors for OLS.
- However, if we know something about the specific form of the heteroskedasticity, we can obtain more efficient estimates than OLS.
- The basic idea is going to be to transform the model into one that has homoskedastic errors – called weighted least squares.

Let $\mathbf{x} = (x_1, x_2, ..., x_k)$ and consider the linear regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u,$$
$$E(u|\mathbf{x}) = 0.$$

- Suppose the heteroskedasticity can be modeled as $Var(u|\mathbf{x}) = \sigma^2 h(\mathbf{x})$.
- **Example:**

  $$Wage = \beta_0 + \beta_1 Education + \beta_2 Experience + \beta_3 Tenure + u$$

  where $Var(u|Education, Experience, Tenure) = \sigma^2 \exp(Education)$.
- $E(u/\sqrt{h(\mathbf{x})}|\mathbf{x}) = 0$, because $h(\mathbf{x})$ is only a function of $\mathbf{x}$, and $Var(u/\sqrt{h(\mathbf{x})}|\mathbf{x}) = \sigma^2$.

- So, if we divide our whole regression equation by $\sqrt{h(\mathbf{x})}$ we have a model where the error is homoskedastic.

$$y^* = \beta_0 x_0^* + \beta_1 x_1^* + \beta_2 x_2^* + ... + \beta_k x_k^* + u^*$$

where

$$\begin{aligned} y^* &= y/\sqrt{h(\mathbf{x})} \\ x_0 &= 1/\sqrt{h(\mathbf{x})} \\ x_j^* &= x_j/\sqrt{h(\mathbf{x})}, j = 1, ..., k \\ u^* &= u/\sqrt{h(\mathbf{x})} \end{aligned}$$

as $E[u^*|\mathbf{x}] = 0$ and $var[u^*|\mathbf{x}] = \sigma^2$.

- Estimating the transformed equation by OLS leads to the *generalized least squares (GLS) estimator*.

- GLS will be *BLUE* in this case, while OLS is *not efficient*.
- The GLS estimator for the particular case where we divide the regression equation by $h(\mathbf{x}_i)$ is called a *weighted least squares (WLS)* estimator. Why?

$\sum_{i=1}^{n}(y_i^* - \hat{\beta}_0/\sqrt{h(\mathbf{x}_i)} - \hat{\beta}_1 x_{i1}^* - ... - \hat{\beta}_k x_{ik}^*)^2$,where $y_i^* = y_i/\sqrt{h(\mathbf{x}_i)}$ and $x_{ij}^* = x_{ij}/\sqrt{h(\mathbf{x}_i)}$

$= \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - ... - \hat{\beta}_k x_{ik})^2/h(\mathbf{x}_i)$.

Individuals with larger variance are given a smaller weight

- WLS is great if we know what $Var(u_i|\mathbf{x}_i)$ looks like.
- In most cases, won't know form of heteroskedasticity.

- We need to estimate $\sigma^2 h(\mathbf{x}_i)$.
- Typically, we start with the assumption of a fairly flexible model, such as $Var(u|\mathbf{x}) = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k)$.
- **Example:**

$$Wage = \beta_0 + \beta_1 Education + \beta_2 Experience + \beta_3 Tenure + u$$

  where
  $Var(u|Education, Experience, Tenure) = \sigma^2 \exp(\delta_0 + \delta_1 Education)$.
- Since we don't know the $\delta' s$, we must estimate them.

- Our assumption implies that $u^2 = \sigma^2 exp(\delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k)v$, where $E(v|\mathbf{x}) = 1$.
- We assume further that v is independent of $\mathbf{x}$.
- Taking logs we obtain

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

  where $E(e) = 0$ and e is independent of x.($\alpha_0 = \log(\sigma^2) + \delta_0 + E(\log(v))$ and $e = \log(v) - E(\log(v))$).
- Now, we know that the residuals $\hat{u}$ is an estimate of u, so if we replace u by $\hat{u}$, we can estimate this equation by OLS. That is, we run the regression of $\log(\hat{u}^2)$ on an intercept x_1, x_2, ..., x_k.
- Denote the fitted values for the observation i by $\widehat{\log(\hat{u}_i^2)}$.

- Now, an estimate of $\sigma^2 h(\mathbf{x}_i)$ is just $\exp\left(\widehat{\log(\hat{u}_i^2)}\right)$, and the inverse of this is our weight.

So, what did we do?

- Run the original OLS model, save the residuals, $\hat{u}$, square them and take the log .
- Regress $\log(\hat{u}^2)$ on all of the independent variables and get the fitted values, $\widehat{\log(\hat{u}_i^2)}$.
- Do WLS using $1/\exp\left(\widehat{\log(\hat{u}_i^2)}\right)$, $i = 1, ..., n$ as weights.

Example: Financial Wealth
We would like to explain the net total financial wealth (*nettfa*)
Observations: 9275.
Regressors
$e401k = 1$ if eligible for 401(k) (pension plan for people in US)
$inc =$annual family income, \$1000s
$male = 1$ if male respondent
$(age - 25)^2$ where age in years
Data set: 1991 US Survey of Income and Program Participation (SIPP).

Example: Financial Wealth

Dependent Variable: *nettfa*

Independent Variables	(1) OLS	(2) WLS	(3) OLS	(4) WLS
inc	.821 (.104)	.787 (.063)	.771 (.100)	.740 (.064)
$(age - 25)^2$	—	—	.0251 (.0043)	.0175 (.0019)
male	—	—	2.48 (2.06)	1.84 (1.56)
e401k	—	—	6.89 (2.29)	5.19 (1.70)
intercept	−10.57 (2.53)	−9.58 (1.65)	−20.98 (3.50)	−16.70 (1.96)
Observations	2,017	2,017	2,017	2,017
R-squared	.0827	.0709	.1279	.1115

- When doing F tests with WLS, form the weights from the unrestricted model and use those weights to do WLS on the restricted model as well as the unrestricted model.
- Remember we are using WLS just for efficiency – OLS is still unbiased & consistent.
- If the Heteroskedastic function is not correct and we estimated the parameters using WLS, we have to use robust standard errors to test hypothesis on the parameters.
- If the Heteroskedastic function is not correct it is not guaranteed that WLS is more efficient than OLS.

1- Suppose that we want an estimate of

$$E(y|x_1 = x_{1,0}, \ldots, x_k = x_{k,0}) = \beta_0 + \beta_1 x_{1,0} + \ldots + \beta_k x_{k,0} = \theta.$$

That is, we would like to estimate the the mean of y when the regressors are equal to known values $x_{1,0}, ..., x_{k,0}$.

- This is easy to obtain by substituting the x's in our estimated model with x_0's ,

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_{1,0} + \ldots + \hat{\beta}_k x_{k,0}.$$

- We would like to construct confidence intervals for θ.
- But what about a standard error of $\hat{y}_0$ under heteroskedasticity?
- θ is just a linear combination of the parameters.

- Can rewrite

$$\beta_0 + \beta_1 x_{1,0} + \ldots + \beta_k x_{k,0} = \theta$$

  as

$$\beta_0 = \theta - \beta_1 x_{1,0} - \ldots - \beta_k x_{k,0}$$

- Substitute in

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

  to obtain

$$y = \theta + \beta_1 (x_1 - x_{1,0}) + \ldots + \beta_k (x_k - x_{k,0}) + u$$

- So, if you regress y on $(x_j - x_{j,0})$, $j = 1, ..., k$, the intercept will give the predicted value. The robust standard errors of the intercept correspond to the standard errors of the prediction under heteroskedasticity.

2- Suppose now that we would like to construct a confidence interval for y when when the regressors are equal to known values $\mathbf{x}_0 = (x_{1,0}, ..., x_{k,0})$ and denote this value as y_0.

- How can we construct a confidence interval for y_0?
- Notice that

$$y_0 = \beta_0 + \beta_1 x_{1,0} + \ldots + \beta_k x_{k,0} + u_0$$

- Our best prediction for y_0 is the regression line

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_{1,0} + \ldots + \hat{\beta}_k x_{k,0}$$

- The prediction error is given by

$$\begin{aligned} \hat{u}_0 &= y_0 - \hat{y}_0 \\ &= \beta_0 + \beta_1 x_{1,0} + \ldots + \beta_k x_{k,0} + u_0 - \hat{y}_0 \end{aligned}$$

- Therefore, The variance of $\hat{u}_0$ conditional on the in-sample values of the independent variables is:

$$\begin{aligned} Var(\hat{u}_0) &= Var(u_0) + Var(\hat{y}_0) \\ &= \sigma^2 h(\mathbf{x}_0) + Var(\hat{y}_0). \end{aligned}$$

$$Var(\hat{u}_0) = \sigma^2 h(\mathbf{x}_0) + Var(\hat{y}_0).$$

- Hence an estimator for $Var(\hat{u}_0)$ is given by

$$se_0^2 = \exp\left(\widehat{\log(\hat{u}_0^2)}\right) + se(\hat{y}_0)^2,$$

where $se(\hat{y}_0)$ is the robust standard error of the intercept in the regression of y on $(x_j - x_{j,0})$, $j = 1, ..., k$, and $\exp\left(\widehat{\log(\hat{u}_0^2)}\right)$ is an estimator of $\sigma^2 h(\mathbf{x}_0)$, computed as in the case of Feasable WLS.
- It can be shown that if $u \sim N(0, \sigma^2 h(\mathbf{x}_0))$,

$$\frac{y_0 - \hat{y}_0}{se_0} \overset{a}{\sim} N(0,1)$$

- Hence the $(1-\alpha)\%$ prediction interval for y_0 is given by

$$(\hat{y}_0 - z_{\alpha/2} se_0, \hat{y}_0 + z_{\alpha/2} se_0),$$

where $z_{\alpha/2}$ is the percentile $(1-\alpha/2)^{th}$ of the standard normal distribution.

Suppose that we have the model

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

$E\left(u|x_1, ..., x_k\right) = 0$, $Var\left(u|x_1, ..., x_k\right) = \sigma^2 h(x_1, ..., x_k)$ and we would like to estimate the the mean of y when the regressors are equal to known values $x_{1,0}, ..., x_{k,0}$: $E(y|x_1 = x_{1,0}, \ldots, x_k = x_{k,0})$.
What can we do?
Given the OLS estimators the predicted value for the mean of $\log(y)$ for any values of the regressors is

$$\widehat{\log(y)} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k$$

If $u \sim N(0, \sigma^2 h(x_1, ..., x_k))$, in can be shown that

$$E\left(y|x_1, ..., x_k\right) = \exp(0.5\sigma^2 h(x_1, ..., x_k)) \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k).$$

Therefore, a simple way to predict $E(y|x_1 = x_{1,0}, \ldots, x_k = x_{k,0})$ is

$$\hat{y}_0 = \exp\left(0.5 \exp\left(\widehat{\log(\hat{u}_0^2)}\right)\right) \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{1,0} + \ldots + \hat{\beta}_k x_{k,0})$$