# Ch08. Heteroskedasticity (Sections 8.1-8.4)

Ping Yu

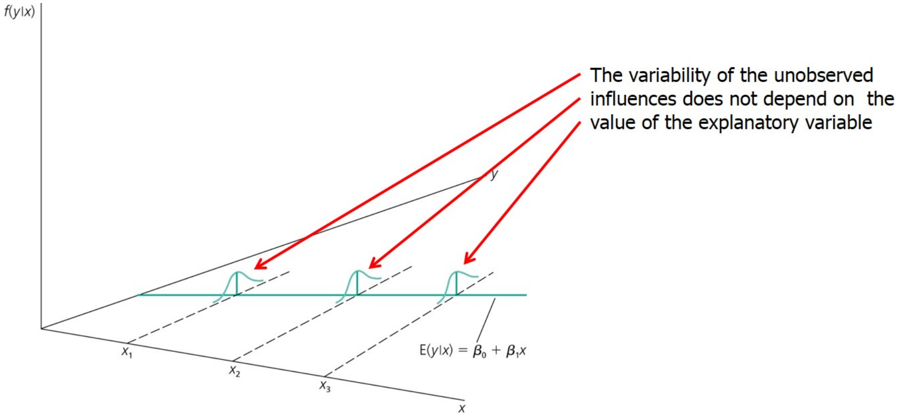
HKU Business School
The University of Hong Kong

# Consequences of Heteroskedasticity for OLS

# Definition of Heteroskedasticity

- If
$$Var\left(u_i|\mathbf{x}_i\right)=\sigma^2$$
is constant, that is, if the variance of the conditional distribution of $u_i$ given $\mathbf{x}_i$ does not depend on $\mathbf{x}_i$, then $u_i$ is said to be homoskedastic. [figure here]
- Otherwise, if
$$Var\left(u_i|\mathbf{x}_i\right)=\sigma^2\left(\mathbf{x}_i\right)\equiv\sigma_i^2,$$
that is, the variance of the conditional distribution of $u_i$ given $\mathbf{x}_i$ depends on $\mathbf{x}_i$, then $u_i$ is said to be heteroskedastic. [figure here]
- Recall that
$$Var\left(u_i|\mathbf{x}_i\right)=E\left[u_i^2|\mathbf{x}_i\right]-E\left[u_i|\mathbf{x}_i\right]^2=E\left[u_i^2|\mathbf{x}_i\right]$$
because $E\left[u_i|\mathbf{x}_i\right]=0$ (Assumption MLR.4).

# Homoskedasticity



# Heteroskedasticity

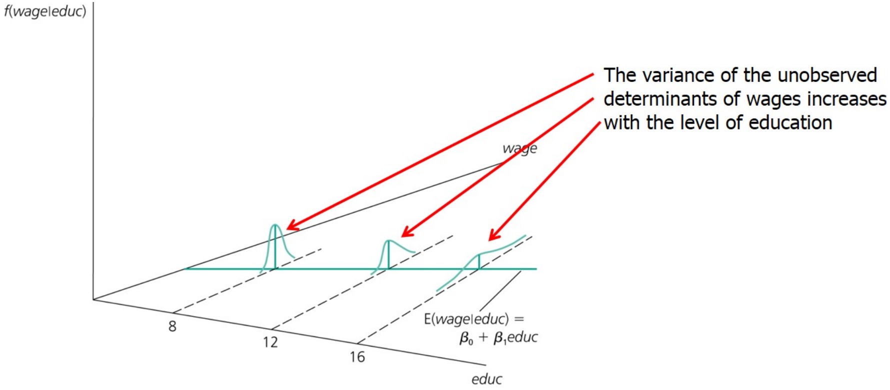



Figure: An Example for Heteroskedasticity: Wage and Education

# A Real-Data Example of Heteroskedasticity

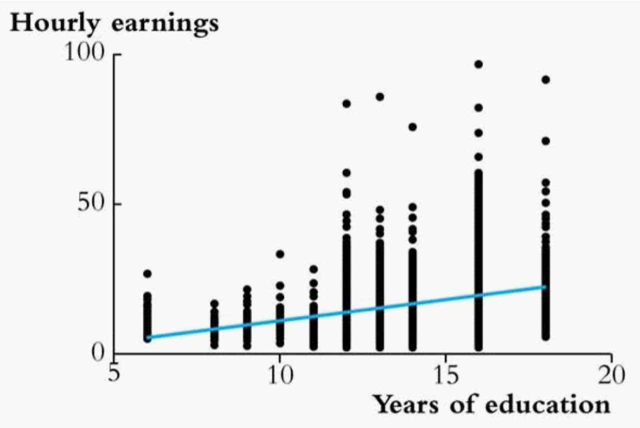



Figure: Average Hourly Earnings vs. Years of Education (data source: Current Population Survey)

# Consequences of Heteroskedasticity for OLS

- OLS is still unbiased under heteroskedasticity because Assumption MLR.4 $E[u|\mathbf{x}] = 0$ does not involve conditional variance.
- Also, interpretation of $R$-squared and $R$-bar squared is not changed because

$$R^2 \approx 1 - \frac{\sigma_u^2}{\sigma_y^2},$$

where $\sigma_u^2$ is the unconditional variance of $u$ while heteroskedasticity is about the conditional variance of $u$.

- Heteroskedasticity invalidates variance formulas for OLS estimators.
- The usual $F$-tests and $t$-tests are not valid under heteroskedasticity because as mentioned before, normality assumption implies homoskedasticity.
- Under heteroskedasticity, OLS is no longer the best linear unbiased estimator (BLUE); there may be more efficient linear estimators as will be discussed below.

# Heteroskedasticity-Robust Inference after OLS Estimation

# Heteroskedasticity-Robust Inference after OLS Estimation

- Formulas for OLS standard errors and related statistics have been developed that are robust to heteroskedasticity of unknown form.
- All formulas are only valid in large samples. (related to chapter 5, which is not discussed in this course)
- Formula for heteroskedasticity-robust OLS standard error:

$$\widehat{Var}\left(\widehat{\beta}_j\right) = \frac{\sum_{i=1}^{n} \widehat{r}_{ij}^2 \widehat{u}_i^2}{SSR_j^2} = SSR_j^{-1} \left[\sum_{i=1}^{n} \widehat{r}_{ij}^2 \widehat{u}_i^2\right] SSR_j^{-1}.$$

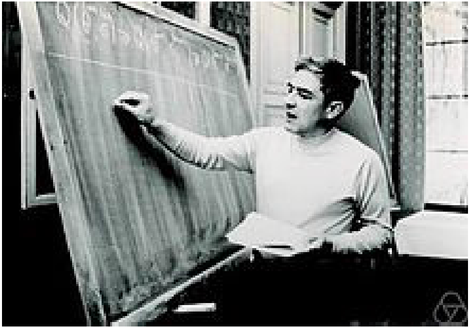
  - Also called Eicker/Huber/White standard errors [photo here] or sandwich-form standard errors. They involve the squared residuals from the regression ($\widehat{u}_i$) and from a regression of $x_j$ on all other explanatory variables ($\widehat{r}_{ij}$). [see the next slides for more discussions]

- Using these formulas, the usual $t$-test is valid asymptotically (i.e., $n \to \infty$).
- The usual $F$-statistic does not work under heteroskedasticity, but heteroskedasticity-robust versions are available in most software.

# Eicker/Huber/White Standard Errors

- This form of standard errors are originally derived in the following two papers:
  - Eicker, F., 1967, Limit Theorems for Regressions with Unequal and Dependent Errors, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 59-82, Berkeley: University of California Press.
  - Huber, P.J., 1967, The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 221-233, Berkeley: University of California Press.
- Later, Halbert White (we will talk more about him later in this chapter) in UCSD derived it independently:
  - White, H., 1980, A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*, 48, 817-838.
- White (1980) is the most-cited paper in economics since 1970. (why?)

Friedhelm Eicker (1927-),
German, Dortmund, 1956MainzPhD



Peter J. Huber (1934-),
Swiss, Bayreuth, 1962ETH ZurichPhD

# (*) Derivation of $Var\left(\widehat{\beta}_j\right)$ Under Heteroskedasticity

- Consider the SLR case and condition on $\{x_i, i = 1, \cdots, n\}$. Recall that

$$\widehat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^n (x_i - \overline{x}) u_i}{\sum_{i=1}^n (x_i - \overline{x})^2}.$$

- If $Var(u_i) = \sigma_i^2$, then

$$Var\left(\widehat{\beta}_1\right) = Var\left(\frac{\sum_{i=1}^n (x_i - \overline{x}) u_i}{\sum_{i=1}^n (x_i - \overline{x})^2}\right) = \frac{\sum_{i=1}^n (x_i - \overline{x})^2 Var(u_i)}{\left[\sum_{i=1}^n (x_i - \overline{x})^2\right]^2} = \frac{\sum_{i=1}^n (x_i - \overline{x})^2 \sigma_i^2}{SST_x^2}.$$

- In the MLR case,

$$Var\left(\widehat{\beta}_j\right) = \frac{\sum_{i=1}^n \widehat{r}_{ij}^2 \sigma_i^2}{SSR_j^2},$$

i.e., replaces $x_i - \overline{x}$ by $\widehat{r}_{ij}$, where $SSR_j = SST_j\left(1 - R_j^2\right) = \sum_{i=1}^n \widehat{r}_{ij}^2$ with $\widehat{r}_{ij}$ being the residual in the regression

$$x_j \sim 1, x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_k.$$

## continue

- In the homoskedastic case, $\sigma_i^2 = \sigma^2$, so

$$Var\left(\widehat{\beta}_j\right) = \frac{\sum_{i=1}^n \widehat{r}_{ij}^2 \sigma^2}{SSR_j^2} = \sigma^2 \frac{\sum_{i=1}^n \widehat{r}_{ij}^2}{SSR_j^2} = \sigma^2 \frac{SSR_j}{SSR_j^2} = \frac{\sigma^2}{SSR_j}$$

  as derived in Chapter 3.
- Note also that

$$Var\left(\widehat{\beta}_j\right) = \frac{\sum_{i=1}^n \widehat{r}_{ij}^2 \sigma_i^2}{SSR_j^2} = \frac{1}{SSR_j} \sum_{i=1}^n \frac{\widehat{r}_{ij}^2}{SSR_j} \sigma_i^2 = \frac{1}{SSR_j} \sum_{i=1}^n w_{ij} \sigma_i^2,$$

  where

$$w_{ij} = \frac{\widehat{r}_{ij}^2}{SSR_j} = \frac{\widehat{r}_{ij}^2}{\sum_{i=1}^n \widehat{r}_{ij}^2} \text{ satisfies } w_{ij} \geq 0 \text{ and } \sum_{i=1}^n w_{ij} = 1.$$

- Compared with the homoskedastic case, the heteroskedasticity-robust variance replaces $\sigma^2$ by a weighted average of $\left\{\sigma_i^2 : i = 1, \cdots, n\right\}$.
- The estimated variance of of $\widehat{\beta}_j$ just replaces $\sigma_i^2$ in $Var\left(\widehat{\beta}_j\right)$ by $\widehat{u}_i^2$.

# (*) Why $\sigma_i^2$ Can Be Estimated by $\widehat{u}_i^2$?

- Consider the simple case where $x$ is binary, e.g., $x = 1(\text{college graduate})$, and $Var(u|x=1) = \sigma_1^2 > \sigma_0^2 = Var(u|x=0)$.
- Suppose the first $n_1$ individuals are college graduates, and the remaining $n_0 = n - n_1$ are noncollege graduates.
- Then

$$Var\left(\widehat{\beta}_1\right) = \frac{\sum_{i=1}^n (x_i - \overline{x})^2 \sigma_i^2}{SST_x^2} = \frac{\sum_{i=1}^{n_1}\left(1-\frac{n_1}{n}\right)^2\sigma_1^2 + \sum_{i=n_1+1}^n\left(0-\frac{n_1}{n}\right)^2\sigma_0^2}{\left[\sum_{i=1}^{n_1}\left(1-\frac{n_1}{n}\right)^2 + \sum_{i=n_1+1}^n\left(0-\frac{n_1}{n}\right)^2\right]^2}$$

$$= \frac{n_1\left(1-\frac{n_1}{n}\right)^2\sigma_1^2 + n_0\left(0-\frac{n_1}{n}\right)^2\sigma_0^2}{\left[n_1\left(1-\frac{n_1}{n}\right)^2 + n_0\left(0-\frac{n_1}{n}\right)^2\right]^2}.$$

- We can estimate $\sigma_1^2$ by $\widehat{\sigma}_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}\widehat{u}_i^2$ and $\sigma_0^2$ by $\widehat{\sigma}_0^2 = \frac{1}{n_0}\sum_{i=n_1+1}^n\widehat{u}_i^2$.

## continue

- Substituting $\sigma_1^2$ by $\widehat{\sigma}_1^2$ and $\sigma_0^2$ by $\widehat{\sigma}_0^2$, we have

$$
\begin{aligned}
\widehat{Var}\left(\widehat{\beta}_1\right) &= \frac{n_1\left(1-\frac{n_1}{n}\right)^2\widehat{\sigma}_1^2+n_0\left(0-\frac{n_1}{n}\right)^2\widehat{\sigma}_0^2}{SST_x^2} \\
&= \frac{n_1\left(1-\frac{n_1}{n}\right)^2\frac{1}{n_1}\sum_{i=1}^{n_1}\widehat{u}_i^2+n_0\left(0-\frac{n_1}{n}\right)^2\frac{1}{n_0}\sum_{i=n_1+1}^{n}\widehat{u}_i^2}{SST_x^2} \\
&= \frac{\sum_{i=1}^{n_1}\left(1-\frac{n_1}{n}\right)^2\widehat{u}_i^2+\sum_{i=n_1+1}^{n}\left(0-\frac{n_1}{n}\right)^2\widehat{u}_i^2}{SST_x^2} \\
&= \frac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2\widehat{u}_i^2}{SST_x^2},
\end{aligned}
$$

  just replacing $\sigma_i^2$ by $\widehat{u}_i^2$.
- Roughly speaking, $\widehat{u}_i^2$ contains information about $\sigma_i^2$.

# Example: Hourly Wage Equation

- The fitted regression line is

$$\begin{aligned} \widehat{\log(wage)} &= -.128 + .0904 educ + .0410 exper - 0.0007 exper^2 \\ &\quad (.105) \quad (.0075) \quad\quad (.0052) \quad\quad (.0001) \\ &\quad [.107] \quad [.0078] \quad\quad [.0050] \quad\quad [.0001] \end{aligned}$$

- Heteroskedasticity-robust standard errors may be larger or smaller (why? check slide 27) than their nonrobust counterparts. The differences are small in this example, but can be quite large if there is strong heteroskedasticity.
  - In most empirical applications, the heteroskedasticity-robust standard errors tend to be larger than the homoskedasticity-only standard errors. In other words, the *t* statistic using the heteroskedasticity-robust standard errors tend to be less significant.
- The null hypothesis here is

$$H_0 : \beta_{exper} = \beta_{exper^2} = 0.$$

- The two $F$ statistics are

$$F = 17.95 \text{ and } F_{robust} = 17.99,$$

  which are not too different in this example, but may be quite different in general.
- Suggestion: To be on the safe side, it is advisable to **always** compute robust s.e.'s.

# Testing for Heteroskedasticity

# Method I: The Graphical Method

- Although we can always use the robust standard errors regardless of homo/hetero, it may still be interesting whether there is heteroskedasticity because then OLS may not be the most efficient linear estimator anymore.
- The key idea of all testing methods is that $\sigma_i^2$ can be approximated by $\widehat{u}_i^2$ (this idea has already been used in $\widehat{Var}\left(\widehat{\beta}_j\right)$). Note that $u_i$ is not observable, so $u_i^2$ is not observable.

- The graphical method just plots $\widehat{u}_i^2$ against $x_i$ to check whether there are some patterns. [figure here]
- With homoskedasticity we'll see something like the first graph: no relationship between $\widehat{u}_i^2$ and the explanatory variable (or combination of explanatory variables such as $\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \cdots + \widehat{\beta}_k x_{ik}$ if $k > 1$).
- Alternatively, with heteroskedasticity we'll see patterns like the other graphs: nonconstant variance (approximated by squared residuals).

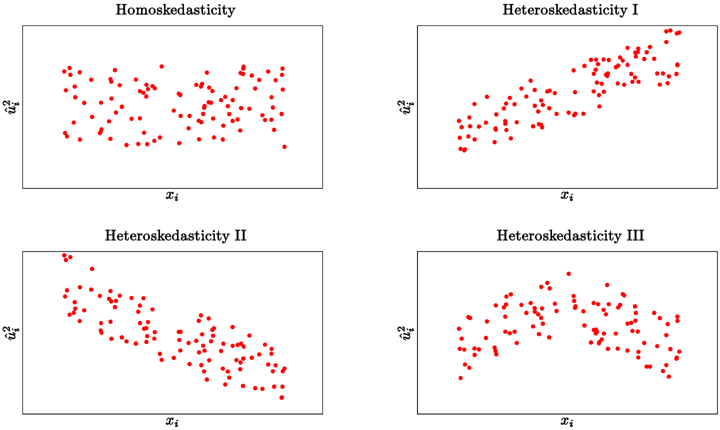



Figure: Graphical Method to Detect Heteroskedasticity

# Method II: The Breusch-Pagan (BP) Test

- Breusch, T.S. and A.R. Pagan, 1979, Simple Test for Heteroskedasticity and Random Coefficient Variation, *Econometrica*, 47, 1287–1294.

Trevor Breusch (1953-),
ANU, 1979ANUPhD

Adrian Pagan (1947-),
University of Sydney, 1972ANUPhD

## continue

- The null hypothesis of the BP test is

$$H_0 : Var\left(u|x_1, \cdots, x_k\right) = Var\left(u|\mathbf{x}\right) = \sigma^2.$$

- Recall that $Var\left(u|\mathbf{x}\right) = E\left[u^2|\mathbf{x}\right]$, so we want to test

$$E\left[u^2|x_1, \cdots, x_k\right] = E\left[u^2\right] = \sigma^2,$$

  that is, the mean of $u^2$ must not vary with $x_1, \cdots, x_k$.

- As a result, we run the following regression:

$$\widehat{u}^2 = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k + error$$

  and test

$$H_0 : \delta_1 = \cdots = \delta_k = 0,$$

  that is, regress squared residuals on all explanatory variables and test whether this regression has explanatory power.

# continue

- The resulting $F$ statistic is (recall the $F$ test for overall significance of a regression in Chapter 4)

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

- A large test statistic (= a high $R$-squared) is evidence against the null hypothesis.
- Alternatively, we can use the Lagrange multiplier (LM) statistic,

$$LM = nR^2_{\hat{u}^2} \sim \chi^2_k.$$

- Again, high $R$-squared leads to rejection of the null hypothesis.
- (*) Why $LM \sim \chi^2_k$? Recall that $F \to \chi^2_k/k$ when $n \to \infty$. While

$$LM = kF \cdot \frac{n}{n-k-1}\left(1 - R^2_{\hat{u}^2}\right),$$

where $\frac{n}{n-k-1} \to 1$ as $n \to \infty$ and $R^2_{\hat{u}^2} \to 0$ under $H_0$.

# Example: Heteroskedasticity in Housing Price Equations

- The fitted regression line is

$$\begin{aligned} \widehat{price} &= \underset{(29.48)}{-21.77} + \underset{(.00064)}{.00207} lotsize + \underset{(.013)}{.123} sqrft + \underset{(9.01)}{13.85} bdrms \\ n &= 88, R^2 = .672 \end{aligned}$$

- The $R$-squared from the regression of $\widehat{u}^2$ on $lotsize, sqrft$ and $bdrms$ is, $R^2_{\widehat{u}^2} = .1601$. The resulting

$$\begin{aligned} F &= \frac{.1601/3}{(1-.1601)/(88-3-1)} \approx 5.34 \text{ with } p\text{-value} = .002, \\ LM &= 88 \times .1601 \approx 14.09 \text{ with } p\text{-value} = .0028. \end{aligned}$$

[figure here]

- So both tests reject the null, and we conclude that the model is heteroskedastic when the dependent variable is *price*.

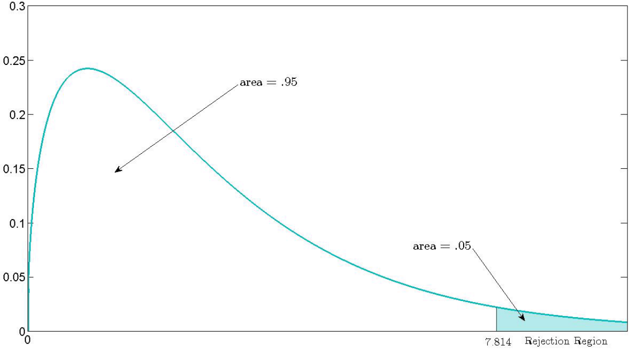



Figure: The 5% Critical Value and Rejection Region in an $\chi_3^2$ Distribution: a $\chi^2$-distributed variable only takes on positive values

# continue

- The fitted regression line after the logarithmic transformation is

$$\begin{aligned} \widehat{\log(price)} &= \underset{(.65)}{-1.30} + \underset{(.038)}{.168}\log(lotsize) + \underset{(.093)}{.700}\log(sqrft) + \underset{(.028)}{.037}bdrms \\ n &= 88, R^2 = .643 \end{aligned}$$

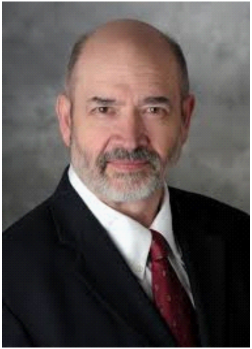
- The $R$-squared from the regression of $\widehat{u}^2$ on $\log(lotsize)$, $\log(sqrft)$ and $bdrms$ is, $R^2_{\widehat{u}^2} = .0480 < .1601$. The resulting

$$\begin{aligned} F &= \frac{.0480/3}{(1-.0480)/(88-3-1)} \approx 1.41 \text{ with } p\text{-value} = .245, \\ LM &= 88 \times .0480 \approx 4.22 \text{ with } p\text{-value} = .239. \end{aligned}$$

- So neither test can reject the null, and we conclude that the model is homoskedastic when the dependent variable is $\log(price)$.
- As mentioned in Chapter 6, taking logs on the dependent variable often helps to secure homoskedasticity.

# Method III: The White Test

- H. White, 1980, A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*, 48, 817-838.



Halbert White, Jr. (1950-2012), UCSD, 1976MITPhD

- Website of his company: http://www.bateswhite.com/

# continue

- Similar to the BP test, but regress squared residuals on all explanatory variables, their squares, and interactions:

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

- Here, $k=3$ results in 9 regressors. Generally, we have $k+k+C_k^2=2k+\frac{k(k-1)}{2}=\frac{k(k+3)}{2}$ regressors.
- The null hypothesis is

$$H_0:\delta_1=\cdots=\delta_9=0,$$

  i.e., the White test detects more general deviations (not only linear form but quadratic form in $\mathbf{x}_i$) from homoskedasticity than the BP test.
- We can apply the $F$ or LM test to detect heteroskedasticity, e.g., $LM=nR^2_{\widehat{u}^2}\sim\chi^2_9$ here.
- (*) Why quadratic form in $\mathbf{x}_i$? This is not because White Taylor expands $\sigma^2(\mathbf{x}_i)$ to the second order rather than the first order as BP. Actually, White tried to test

$$\widehat{Var}\left(\widehat{\beta}_j\right)=\frac{\sum_{i=1}^n \frac{\widehat{r}_{ij}^2}{SSR_j}\widehat{u}_i^2}{SSR_j}\overset{?}{=}\frac{\frac{1}{n}\sum_{i=1}^n \widehat{u}_i^2}{SSR_j}=\frac{\widetilde{\sigma}^2}{SSR_j}=\widehat{Var}_{\text{homo}}\left(\widehat{\beta}_j\right);$$

  this is why the quadratic terms of $\mathbf{x}_i$ appear.

# Disadvantage of This Form of the White Test

- Including all squares and interactions leads to a large number of estimated parameters. E.g. $k = 6$ leads to 27 parameters to be estimated.
- It is suggested to use an alternative form of the White test:

$$\widehat{u}^2 = \delta_0 + \delta_1 \widehat{y} + \delta_2 \widehat{y}^2 + error,$$

  where $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \cdots + \widehat{\beta}_k x_k$ is the predicted value, and $\delta_1 \widehat{y} + \delta_2 \widehat{y}^2$ is a **special** (why?) quadratic form of $\mathbf{x}_i$.
- The null hypothesis here is

$$H_0 : \delta_1 = \delta_2 = 0,$$

  and the $F$ or LM statistics can apply, e.g., $LM = nR^2_{\widehat{u}^2} \sim \chi^2_2$.
- Example (Heteroskedasticity in Log Housing Price Equations): We regress $\widehat{u}^2$ on $\widehat{lprice}$ and $\widehat{lprice}^2$ in the above example. $R^2_{\widehat{u}^2} = .0392$, so

$$LM = 88 \times .0392 \approx 3.45 \text{ with } p\text{-value} = .178 < .239,$$

  but we still cannot reject the model is homoskedastic as the BP test.

# Weighted Least Squares Estimation

# a: The Heteroskedasticity Is Known up to a Multiplicative Constant

- Suppose
$$Var(u|\mathbf{x}) = \sigma^2 h(\mathbf{x}),$$
where $h(\mathbf{x}) > 0$ is known, but $\sigma^2$ is unknown.
- Note that
$$\sigma_i^2 = \sigma^2 h(\mathbf{x}_i) \equiv \sigma^2 h_i.$$
- In the regression,
$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i,$$
if we transform the model by dividing both sides by $\sqrt{h_i}$, then we have
$$\frac{y_i}{\sqrt{h_i}} = \beta_0 \frac{1}{\sqrt{h_i}} + \beta_1 \frac{x_{i1}}{\sqrt{h_i}} + \cdots + \beta_k \frac{x_{ik}}{\sqrt{h_i}} + \frac{u_i}{\sqrt{h_i}},$$
denoted as
$$y_i^* = \beta_0 x_{i0}^* + \beta_1 x_{i1}^* + \cdots + \beta_k x_{ik}^* + u_i^*,$$
where $x_{i0}^* = \frac{1}{\sqrt{h_i}}$, i.e., this regression model has no intercept.

# The Transformed Model Is Homoskedastic

- Note that
$$E\left[u_i^*|\mathbf{x}_i\right]=E\left[\frac{u_i}{\sqrt{h_i}}\middle|\mathbf{x}_i\right]=\frac{1}{\sqrt{h_i}}E\left[u_i|\mathbf{x}_i\right]\overset{MLR.4}{=}0,$$
so
$$\begin{aligned}Var\left(u_i^*|\mathbf{x}_i\right) &= E\left[u_i^{*2}|\mathbf{x}_i\right]-E\left[u_i^*|\mathbf{x}_i\right]^2=E\left[u_i^{*2}|\mathbf{x}_i\right]\\ &= E\left[\left(\frac{u_i}{\sqrt{h_i}}\right)^2\middle|\mathbf{x}_i\right]=E\left[\frac{u_i^2}{h_i}\middle|\mathbf{x}_i\right]\\ &= \frac{1}{h_i}E\left[u_i^2\middle|\mathbf{x}_i\right]=\frac{1}{h_i}\sigma^2h_i=\sigma^2.\end{aligned}$$

- Example (Savings and Income): Consider the simple savings function,
$$sav_i=\beta_0+\beta_1 inc_i+u_i, Var\left(u_i|inc_i\right)=\sigma^2 inc_i.$$ [figure here]

  After the transformation,
$$\frac{sav_i}{\sqrt{inc_i}}=\beta_0\frac{1}{\sqrt{inc_i}}+\beta_1\frac{inc_i}{\sqrt{inc_i}}+u_i^*,$$
  where $u_i^*$ is homoskedastic - $Var\left(u_i^*|inc_i\right)=\sigma^2$.

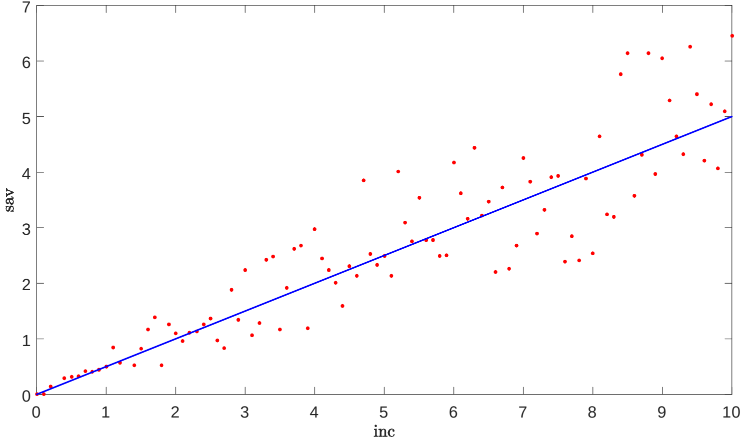



Figure: $sav_i = \beta_0 + \beta_1 inc_i + u_i$ with $\beta_0 = 0$, $\beta_1 = 0.5$ and $Var(u_i|inc_i) = 0.09 inc_i$

# Weighted Least Squares (WLS)

- OLS in the transformed model is weighted least squares (WLS):

$$\min_{\beta_0,\beta_1,\cdots,\beta_k} \sum_{i=1}^{n} \left( \frac{y_i}{\sqrt{h_i}} - \beta_0 \frac{1}{\sqrt{h_i}} - \beta_1 \frac{x_{i1}}{\sqrt{h_i}} - \cdots - \beta_k \frac{x_{ik}}{\sqrt{h_i}} \right)^2$$

$$\Longleftrightarrow \quad \min_{\beta_0,\beta_1,\cdots,\beta_k} \sum_{i=1}^{n} \frac{1}{h_i} \left( y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik} \right)^2.$$

- Observations with a large variance get a smaller weight in the optimization problem because $h_i \propto \sigma_i^2$.
- Why is WLS more efficient than OLS in the original model? Observations with a large variance are less informative than observations with small variance and therefore should get less weight. [check the saving example again]
- If the other Gauss-Markov assumptions hold as well, OLS applied to the transformed model, or the WLS, is the best linear unbiased estimator (BLUE).
- WLS is a special case of generalized least squares (GLS) which accounts for **also** serial correlation among $\{u_i, i = 1, \cdots, n\}$.

# (*) Which $R$-Squared Is Calculated?

- Be cautious about which $R^2$ is reported for WLS in practice.
- There are two methods: recall that

$$R^2 = 1 - \frac{SSR}{SST}.$$

- (I): The dependent variable is $y_i$ (i.e., $SST = \sum_{i=1}^n (y_i - \overline{y})^2$), and $SSR$ is calculated from

$$\widehat{u}_i = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_{i1} - \cdots - \widehat{\beta}_k x_{ik}.$$

  - In this case, $R^2$ for WLS is smaller than OLS although WLS is more efficient than OLS. This is because they share the same $SST$, but OLS minimizes $SSR$.
  - This should be the correct $R^2$ to be reported.

- (II): The dependent variable is $\frac{y_i}{\sqrt{h_i}}$ (i.e., $SST = \sum_{i=1}^n \left(\frac{y_i}{\sqrt{h_i}} - \overline{\frac{y}{\sqrt{h}}}\right)^2$. This is not correct because the transformed model does not include an intercept which is required for $R^2$ calculation), and $SSR$ is calculated from

$$\widetilde{u}_i = \frac{y_i}{\sqrt{h_i}} - \widehat{\beta}_0 \frac{1}{\sqrt{h_i}} - \widehat{\beta}_1 \frac{x_{i1}}{\sqrt{h_i}} - \cdots - \widehat{\beta}_k \frac{x_{ik}}{\sqrt{h_i}}.$$

  - Quite often, people just report the $R^2$ from the transformed model carelessly, e.g., copy from standard econometric softwares.

- Check the following example.

# Example: Financial Wealth Equation

Net financial wealth

**TABLE 8.1 Dependent Variable: *nettfa***

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



Assumed form of heteroscedasticity:
$Var(u|inc, age, male, e401k) = \sigma^2 inc$

WLS estimates have considerably smaller standard errors (which is in line with the expectation that they are more efficient).

Participation in 401K pension plan

- $R^2_{OLS} > R^2_{WLS}$, which is also possible when method (II) of the last slide is used to calculate $R^2_{WLS}$, so we cannot tell which method of $R^2_{WLS}$ is used in this example [see slide 42 for a counter-example].

# 401(k) Plan

- In the early 1980s, the United States introduced several tax-deferred savings options designed to increase individual savings for retirement, the most popular being Individual Retirement Accounts (IRAs) and 401(k) plans.
- IRAs and 401(k) plans are similar in that both allow the individual to deduct contributions to retirement accounts from taxable income and they both permit tax-free accrual of interest.
- The key difference between these schemes is that employers provide 401(k) plans and may match some percentage of the employee 401(k) contributions.
- Therefore, only workers in firms that offer such programs are eligible, whereas IRAs are open to all.

# Important Special Case of Heteroskedasticity

- If the observations are reported as averages at the city/county/state/country/firm level, they should be weighted by the size of the unit.
- Suppose we have only average values related to 401(k) contributions at the firm level, but the individual level data are not available.
- Suppose the individual level equation is

$$contrib_{i,e} = \beta_0 + \beta_1 earns_{i,e} + \beta_2 age_{i,e} + \beta_3 mrate_i + u_{i,e},$$

where
$contrib_{i,e} =$ annual contribution by employee $e$ who works for firm $i$
$earns_{i,e} =$ annual earning for this person
$mrate_i =$ the amount the firm puts into an employee's account for every dollar the employee contributes or the match rate

- Then the firm level equation is

$$\overline{contrib}_i = \beta_0 + \beta_1 \overline{earns}_i + \beta_2 \overline{age}_i + \beta_3 mrate_i + \overline{u}_i,$$

where the error term $\overline{u}_i = \frac{1}{m_i} \sum_{e=1}^{m_i} u_{i,e}$ has variance [check Chapter 2, slide 67]

$$Var\left(\overline{u}_i\right) = \frac{\sigma^2}{m_i}$$

if $Var\left(u_{i,e}\right) = \sigma^2$, i.e., $u_{i,e}$ is homoskedastic at the employee level, where $m_i$ is the number of employees at firm $i$.

# continue

- $h_i = \frac{1}{m_i}$, so in the WLS,

$$\begin{aligned} & \min_{\beta_0,\beta_1,\cdots,\beta_k} \sum_{i=1}^{n} \frac{1}{h_i} \left( y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik} \right)^2 \\ = & \min_{\beta_0,\beta_1,\cdots,\beta_k} \sum_{i=1}^{n} m_i \left( y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik} \right)^2 . \end{aligned}$$

- In summary, if errors are homoskedastic at the employee level, WLS with weights equal to firm size $m_i$ should be used.
- If the assumption of homoskedasticity at the employee level is not exactly right, one can calculate robust standard errors after WLS (i.e., for the transformed model). [see more discussion later]
- That is, it is always a good idea to compute fully robust standard errors and test statistics after WLS estimation.

# (*) b: Unknown Heteroskedasticity Function (Feasible GLS)

- Assume
$$Var(u|\mathbf{x}) = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k) = \sigma^2 h(\mathbf{x}),$$
where exp-function is used to ensure positivity.
- Under this assumption, we can write
$$u^2 = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k) v,$$
where $v$ is a multiplicative error independent of the explanatory variables, and $E[v] = 1$ (why?). (why $v$ appears? why multiplicative $v$?)
- Now,
$$\begin{aligned} \log u^2 &= \log\left(\sigma^2\right) + \delta_0 + E[\log v] + \delta_1 x_1 + \cdots + \delta_k x_k + \log v - E[\log v] \\ &\equiv \alpha_0 + \delta_1 x_1 + \cdots + \delta_k x_k + e, \end{aligned}$$
where $e = \log v - E[\log v]$ with $E[e] = 0$, and $\alpha_0 = \log\left(\sigma^2\right) + \delta_0 + E[\log v]$.
- (**) From slides 36 and 37 of Chapter 6, $E[\log v] \leq \log(E[v]) = \log 1 = 0$, so we need to demean $\log v$ to guarantee $E[e] = 0$.

# continue

- Regress $\log\left(\widehat{u}^2\right)$ on $1, x_1, \cdots, x_k$ to have $\left(\widehat{\alpha}_0, \widehat{\delta}_1, \cdots, \widehat{\delta}_k\right)$; then the estimated $h(\mathbf{x})$ is
$$\widehat{h}(\mathbf{x}) = \exp\left(\widehat{\alpha}_0 + \widehat{\delta}_1 x_1 + \cdots + \widehat{\delta}_k x_k\right),$$
where the term $\widehat{\alpha}_0$ can be neglected because it can be absorbed in the constant $\sigma^2$.
- As in the White test, we can regress $\log\left(\widehat{u}^2\right)$ on $1, \widehat{y}$ and $\widehat{y}^2$ to estimate $h(\mathbf{x})$.
- Caution: In the BP or White test, the dependent variable is $\widehat{u}^2$ rather than $\log\left(\widehat{u}^2\right)$! Otherwise, the distribution of the $F$ statistic is more complicated and stronger assumptions (e.g., $u$ and $\mathbf{x}$ are independent) are required.
- Using inverse values of the estimated heteroskedasticity function as weights in WLS, we get the feasible GLS (FGLS) estimator.
- Feasible GLS is consistent and asymptotically more efficient than OLS if the assumption on the form of $Var\left(u|\mathbf{x}\right)$ is correct (if incorrect, discuss later).

# (*) Example: Demand for Cigarettes

- The estimated demand equation for cigarettes by OLS is

$$
\begin{aligned}
\widehat{cigs} &= -3.64 + .880 \log(income) - .75 \log(cigpric) \\
&\quad (24.08)(.728) \qquad\qquad\qquad (5.773) \\
&\quad -.501 educ + .771 age - .0090 age^2 - 2.83 restaurn \\
&\quad (.167) \qquad (.160) \qquad (.0017) \qquad (1.11) \\
n &= 807, R^2 = .0526 \text{ (quite small)}
\end{aligned}
$$

  where
  $cigs$ = number of cigarettes smoked per day
  $income$ = annual income
  $cigpric$ = the per-pack price of cigarettes (in cents)
  $restaurn$ = a binary indicator for smoking restrictions in restaurants

- The income effect is insignificant.
- The $p$-value for the BP test (either $F$ or LM) is .000, i.e., there is strong evidence that the model is heteroskedastic.

# continue

- If estimated by FGLS, then

$$\begin{aligned}\widehat{cigs} &= \underset{(17.80)}{-5.64}+\underset{(.44)}{1.30}\log(income)-\underset{(4.46)}{2.94}\log(cigpric)\\ &\quad -\underset{(.120)}{.463}educ+\underset{(.097)}{.482}age-\underset{(.0009)}{.0056}age^2-\underset{(.80)}{3.46}restaurn\\ n &= 807, R^2 = .1134\ (> .0526\text{, so method (II) on slide 34 is used})\end{aligned}$$

- The income effect is now statistically significant.
- Other coefficients are also more precisely estimated (without changing qualitative results).
- Interestingly, the turnaround point of age in both models is about 43:

$$\frac{.771}{2\times .0090} \approx \frac{.482}{2\times .0056} \approx 43.$$

# (*) c: What If the Assumed Heteroskedasticity Function is Wrong?

- If the heteroskedasticity function is misspecified, WLS is still consistent under MLR.1-MLR.4, but robust standard errors should be computed since

$$Var\left(u_i^*|\mathbf{x}_i\right)=\sigma^2\frac{h_i}{g_i}\neq\sigma^2,$$

  where it is assumed $Var\left(u|\mathbf{x}\right)=\sigma^2 g\left(\mathbf{x}\right)$ with $g\left(\mathbf{x}\right)$ not equal to the true heteroskedasticity $h\left(\mathbf{x}\right)$.
- WLS is consistent under MLR.4 but not necessarily under MLR.4$'$ [WLS vs. OLS: trade-off between efficiency and robustness], where recall that

$$\text{MLR.4: } E\left[u|\mathbf{x}\right]=0\Longrightarrow\text{MLR.4}':\ E\left[\mathbf{x}u\right]=0.$$

- MLR.4 implies

$$E\left[u_i^*|\mathbf{x}_i\right]=E\left[\left.\frac{u_i}{\sqrt{h_i}}\right|\mathbf{x}_i\right]=0,$$

  but MLR.4$'$ does not.
- If OLS and WLS produce very different estimates of $\beta$, this typically indicates that some other assumptions (e.g., MLR.4) are wrong.
- If there is strong heteroskedasticity, it is still often better to use a wrong form of heteroskedasticity in order to increase efficiency if $\frac{h(\mathbf{x})}{g(\mathbf{x})}$ is more constant than $h\left(\mathbf{x}\right)$, i.e., part of heteroskedasticity is offsetted.