# Introductory Econometrics

## Heteroskedasticity

Yaohan Chen

School of Big Data and Statistics, Anhui University

Spring, 2025

# Homoskedasticity and Heteroskedasticity

- Multivariate linear regression model expressed as follows

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + u_i.$$

- Conditional homoskedasticity:

$$\mathrm{E}\left(u_i^2 \mid X_{i1}, X_{i2}, \cdots, X_{ik}\right) = \sigma^2$$

- Conditional heteroskedasticity:

$$\mathrm{E}\left(u_i^2 \mid X_{i1}, X_{i2}, \cdots, X_{ik}\right) = \sigma_i^2 = f(X_{i1}, X_{i2}, \cdots, X_{ik})$$

# Homoskedasticity and Heteroskedasticity

- Recall that under the classical assumptions, we have

$$\mathrm{E}\left(\boldsymbol{u}\boldsymbol{u}' \mid \boldsymbol{X}\right) = \begin{pmatrix} \mathrm{E}\left(u_1^2 \mid \boldsymbol{X}\right) & \cdots & \mathrm{E}\left(u_1 u_n \mid \boldsymbol{X}\right) \\ \vdots & \ddots & \vdots \\ \mathrm{E}\left(u_n u_1 \mid \boldsymbol{X}\right) & \cdots & \mathrm{E}\left(u_n^2 \mid \boldsymbol{X}\right) \end{pmatrix} = \sigma^2 \boldsymbol{I}$$

  which is crucial for proving many theoretical properties.
- From conditional homoskedasticity to conditional heteroskedasticity, $\mathrm{E}\left(\boldsymbol{u}\boldsymbol{u}' \mid \boldsymbol{X}\right)$ is conditional on $\boldsymbol{X}$ in general, i.e. $\sigma^2 \boldsymbol{V}\left(\boldsymbol{X}\right)$. $\boldsymbol{V}\left(\boldsymbol{X}\right)$ is finite positive definite matrix.

# Issues Associated with Heteroskedasticity

- OLS estimator is not efficient.
- Variance of OLS estimator $\hat{\boldsymbol{\beta}}$, $\text{Var}\left(\hat{\boldsymbol{\beta}}\right)$ is not $\sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$. Therefore, hypothesis testing statistic changes.
- For instance, $\hat{\beta}_1$ for the simple linear regression is $\frac{\sum x_i Y_i}{\sum x_i^2}$ and $\text{Var}\left(\hat{\beta}_1\right) = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2}$ when heteroskedasticity arises.
- $\frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}}$ follows student $t$ distribution when conditional homoskedasticity holds, where $S_{\hat{\beta}_1}$ is the estimation of standard deviation of of $\hat{\beta}_1$, i.e. $\sqrt{\text{Var}\left(\hat{\beta}_1\right)}$.

# Issues Associated with Heteroskedasticity

- Besides, for prediction,

$$\frac{\hat{Y}_0 - \mathrm{E}\left(Y_0\right)}{\hat{\sigma}\sqrt{\boldsymbol{X_0}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}_0'}} \sim t(n-k-1)$$

  which is the result used for constructing prediction confidence interval. This result holds when homoskedasticity holds BUT changes when heteroskedasticity arises.
- Heteroskedasticity also affects the prediction of the linear regression models based on OLS.

# Testing for Heteroskedasticity

- Testing for heteroskedasticity can be equivalently understood as testing for homoskedasticity. That is,

$$H_0 : \mathrm{E}\left(u_i^2 \mid \boldsymbol{X}_i\right) = \sigma^2 \text{ versus } H_1\text{: not } H_0.$$

  where $\boldsymbol{X}_i = (1, X_{i1}, X_{i2}, \cdots, X_{ik})'$.

- **Breusch-Pagan LM Test**. For the established multivariate linear regression model, we may consider following auxiliary regression

$$u_i^2 = \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \cdots + \delta_k X_{ik} + \varepsilon_i$$

  and use the $e_i^2$ (squared residuals from the original regression) as the proxy for $u_i^2$.

# Testing for Heteroskedasticity

- Under $H_0$, one should expect

$$H_0 : \delta_1 = \delta_2 = \cdots = \delta_k = 0.$$

- It can be shown (Breusch and Pagan, 1979, ECMA; Koenker and Bassett, 1982, ECMA) that under $H_0$ we have test statistics

$$\begin{aligned} F_n &= \frac{R_{e^2}^2/k}{\left(1 - R_{e^2}^2\right)/(n-k-1)} \sim F\left(k, n-k-1\right) \\ LM_n &= nR_{e^2}^2 \sim \chi^2(k) \end{aligned}$$

where $R_{e^2}^2$ refers to the $R^2$ of the auxiliary regression.

# Testing for Heteroskedasticity (∗)

- Under $H_0 : \mathrm{E}\left(u_i^2 \mid \boldsymbol{X}_i\right) = \sigma^2$, recall the large sample properties of OLS when

$$P \lim \boldsymbol{X}'\boldsymbol{X}/n = \boldsymbol{Q} \quad \text{or} \quad \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i' \xrightarrow{p} \boldsymbol{Q}$$

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N\left(0, \sigma^2 \boldsymbol{Q}^{-1}\right) = N\left(0, \boldsymbol{Q}^{-1} \boxed{\sigma^2 \boldsymbol{Q}} \boldsymbol{Q}^{-1}\right),$$

- But alternatively when heteroskedasticity arises, one should expect

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} N\left(0, \boldsymbol{Q}^{-1} \boxed{\boldsymbol{V}} \boldsymbol{Q}^{-1}\right)$$

where

$$\boldsymbol{V} \equiv \mathrm{E}\left(\boldsymbol{X}_i \boldsymbol{X}_i' u_i^2\right).$$

# Testing for Heteroskedasticity (∗)

- **White's Test** is inspired by comparing $\sigma^2 \boldsymbol{Q}$ and $\boldsymbol{V}$. Under $H_0$, one should expect $\sigma^2 \boldsymbol{Q} = \boldsymbol{V}$.
- Recall that

$$\hat{\sigma}^2 \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i' \xrightarrow{p} \sigma^2 \boldsymbol{Q} \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i' e_i^2 \xrightarrow{p} \boldsymbol{V}$$

where $\hat{\sigma}^2 = \frac{1}{n-k-1} \sum_{i=1}^{n} e_i^2$ and $e_i$ refers to the residual corresponding to multivariate regression model with $k$ non-constant explanatory variables.

# Testing for Heteroskedasticity ($*$)

- Consequently, under $H_0$,

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'e_i^2-\hat{\sigma}^2\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i' = \frac{1}{n}\sum_{i=1}^{n}\left(e_i^2-\hat{\sigma}^2\right)\boldsymbol{X}_i\boldsymbol{X}_i' \xrightarrow{p} 0$$

which is equivalent to the claim that

$$c_n \equiv \frac{1}{n}\sum_{i=1}^{n}\left(e_i^2-\hat{\sigma}^2\right)\psi_i \xrightarrow{p} 0$$

where $\psi_i$ denotes a vector collecting unique and nonconstant elements of $\boldsymbol{X}_i\boldsymbol{X}_i'$.

# Testing for Heteroskedasticity (∗)

- White (1980, ECMA) shows that under some conditions appropriate for a CLT to be applicable, one would expect that

  $$\sqrt{n}c_n \xrightarrow{d} N(0, C)$$

  where $C$ is the asymptotic variance of $\sqrt{n}c_n$. Furthermore,

  $$T_{n,\hat{C}} \equiv nc_n'\hat{C}^{-1}c_n \xrightarrow{d} \chi^2(p),$$

  where $\hat{C}$ is a consistent estimator for $C$. $p$ denotes the dimension of $c_n$.

# Testing for Heteroskedasticity (*)

- White (1980) shows that for a certain choice of $\hat{C}$, $T_{n,\hat{C}}$ can be calculated as $nR^2_{e^2}$, where $R^2_{e^2}$ refers to the $R^2$ of the auxiliary regression of $e_i^2$ on a constant and $\psi_i$, i.e.,

$$T_n = nR^2_{e^2}.$$

- Under $H_0$, we have

$$T_n = nR^2_{e^2} \xrightarrow{d} \chi^2(p).$$

# Testing for Heteroskedasticity

- Suppose $\boldsymbol{X}_i = (1, X_{i1}, X_{i2})'$ and we have following regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + u_i.$$

Then,

$$\begin{pmatrix} 1 & X_{i1} & X_{i2} \\ X_{i1} & X_{i1}^2 & X_{i1}X_{i2} \\ X_{i2} & X_{i1}X_{i2} & X_{i2}^2 \end{pmatrix}.$$

- The auxiliary regression is constructed as

$$e_i^2 = \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 X_{i1}^2 + \delta_4 X_{i2}^2 + \delta_5 X_{i1}X_{i2} + \varepsilon_i.$$

- Reject $H_0$ if $nR^2_{e^2} > \chi^2(p)$.

# Remedies for Heteroskedasticity

- $\mathrm{E}(\boldsymbol{u}\boldsymbol{u}' \mid \boldsymbol{X}) = \sigma^2 \boldsymbol{V}(\boldsymbol{X})$ conditional on $\boldsymbol{X}$ allows for heteroskedasticity.
- For linear regression model under heteroskedasticity

  $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}\ ,\ \mathrm{E}(\boldsymbol{u} \mid \boldsymbol{X}) = 0\ ,\ \mathrm{E}(\boldsymbol{u}\boldsymbol{u}' \mid \boldsymbol{X}) = \sigma^2 \boldsymbol{V}(\boldsymbol{X}).$

  Pre-multiplying both sides by $C$ yields

  $$C\boldsymbol{Y} = C\boldsymbol{X}\boldsymbol{\beta} + C\boldsymbol{u}$$

  and note that $\mathrm{Var}(C\boldsymbol{u}) = \sigma^2 \boldsymbol{I}$ if $[\boldsymbol{V}(\boldsymbol{X})]^{-1} = C'C$.

# Remedies for Heteroskedasticity

- If we define $C\boldsymbol{Y}$ as $\boldsymbol{Y}^*$, $C\boldsymbol{X}$ as $\boldsymbol{X}^*$, and $C\boldsymbol{u}$ as $\boldsymbol{u}^*$, then the transformed linear regression model

$$\boldsymbol{Y}^* = \boldsymbol{X}^*\boldsymbol{\beta} + \boldsymbol{u}^*$$

  satisfy the classical assumptions.

- OLS estimator for $\boldsymbol{\beta}$ is

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^* &= \left(\boldsymbol{X}^{*\prime}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*\prime}\boldsymbol{Y}^* \\
&= \left(\boldsymbol{X}'C'C\boldsymbol{X}\right)^{-1}\boldsymbol{X}'C'C\boldsymbol{Y} \\
&= \left\{\boldsymbol{X}'\left[\boldsymbol{V}(\boldsymbol{X})\right]^{-1}\boldsymbol{X}\right\}^{-1}\boldsymbol{X}'\left[\boldsymbol{V}(\boldsymbol{X})\right]^{-1}\boldsymbol{Y} \\
&\equiv \hat{\boldsymbol{\beta}}_{GLS} \text{ (Generalized Least Square Estimator).}
\end{aligned}$$

# Generalized Least Square Estimator

- **Unbiasedness**. $\mathrm{E}\left(\hat{\boldsymbol{\beta}}_{GLS} \mid \boldsymbol{X}\right) = \boldsymbol{\beta}$.
- **Variance-covariance matrix**.

$$\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_{GLS} \mid \boldsymbol{X}\right) = \sigma^2 \left(\boldsymbol{X}' \left[\boldsymbol{V}(\boldsymbol{X})\right]^{-1} \boldsymbol{X}\right)^{-1}.$$

- $\hat{\boldsymbol{\beta}}_{GLS}$ is BLUE estimator for linear regression model with heteroskedasticity.
- **Unbiasedness of** $\hat{\sigma}^{2*}$. $\mathrm{E}\left(\hat{\sigma}^{2*} \mid \boldsymbol{X}\right) = \sigma^2$, where

$$\hat{\sigma}^{2*} = \frac{1}{n-k-1} \boldsymbol{e}^{*\prime} \boldsymbol{e}^{*}, \ \boldsymbol{e}^{*} = \boldsymbol{Y}^{*} - \boldsymbol{X}^{*} \hat{\boldsymbol{\beta}}_{GLS}.$$

- **Orthogonality**. $\mathrm{E}\left[\left(\hat{\boldsymbol{\beta}}_{GLS} - \boldsymbol{\beta}\right) \boldsymbol{e}^{*\prime}\right] = 0$.

# More about Generalized Least Square Estimator

- In practice, $\boldsymbol{V}(\boldsymbol{X})$ is generally unknown so that $\hat{\boldsymbol{\beta}}_{GLS}$ is usually infeasible. If one can consistently estimate $\widehat{\boldsymbol{V}(\boldsymbol{X})}$, then we can use the feasible GLS estimator (FGLS)

$$\hat{\boldsymbol{\beta}}_{FGLS} \equiv \left( \boldsymbol{X}' \left[ \widehat{\boldsymbol{V}(\boldsymbol{X})} \right]^{-1} \boldsymbol{X} \right)^{-1} \boldsymbol{X}' \left[ \widehat{\boldsymbol{V}(\boldsymbol{X})} \right]^{-1} \boldsymbol{Y}.$$

- GLS is of more theoretical analysis values rather than practical values.

# Heteroskedasticity-Consistent Standard Errors

- Recall that

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\xrightarrow{d}N\left(0,\boldsymbol{Q}^{-1}\boldsymbol{V}\boldsymbol{Q}^{-1}\right)$$

  where $\boldsymbol{Q}^{-1}\boldsymbol{V}\boldsymbol{Q}^{-1}$ denotes the asymptotic variance (variance-covariance) matrix of $\sqrt{n}\hat{\boldsymbol{\beta}}$, i.e.

$$\operatorname{avar}\left(\sqrt{n}\hat{\boldsymbol{\beta}}\right)=\boldsymbol{Q}^{-1}\boldsymbol{V}\boldsymbol{Q}^{-1}.$$

- Heteroskedasticity mainly causes problems through variance. For large sample, we consider the consistent estimator of $\operatorname{avar}\left(\sqrt{n}\hat{\boldsymbol{\beta}}\right)$.

# Heteroskedasticity-Consistent Standard Errors

- White (1980) proves that under the regular conditions, for OLS estimator $\hat{\boldsymbol{\beta}}$,

$$\widehat{\text{avar}}(\sqrt{n}\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{Q}}^{-1}\hat{\boldsymbol{V}}\hat{\boldsymbol{Q}}^{-1} \xrightarrow{p} \boldsymbol{Q}^{-1}\boldsymbol{V}\boldsymbol{Q}^{-1}$$

  where

$$\hat{\boldsymbol{Q}} \equiv \frac{1}{n}\boldsymbol{X}'\boldsymbol{X} \quad \hat{\boldsymbol{V}} \equiv \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'e_i^2.$$

  and $\boldsymbol{X}_i$ as a column vector denote transpose of the $i$th row of $\boldsymbol{X}$.

- $\widehat{\text{avar}}(\sqrt{n}\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{Q}}^{-1}\hat{\boldsymbol{V}}\hat{\boldsymbol{Q}}^{-1}$ is called the heteroskedasticity-consistent standard variance (variance-covariance) matrix estimator.