Econometrics Lecture 10

Heteroskedasticity, GLS and WLS

Hang Miao

Rutgers University

April 13, 2021

Overview

1. Handle Heteroskedasticity: WLS approach
   - Heteroskedasiticity with Known Functional Form
   - Heteroskedasiticity with Unknown Functional Form
2. Handle Heteroskedasticity: GLS approach
   - Heteroskedasiticity and Serial Correlation with Known Functional Form
3. Handle Heteroskedasticity: Robust Standard Error
   - Heteroskedasiticity with Known Functional Form
   - Heteroskedasiticity with Unknown Functional Form
4. Connect to the Assumptions in Textbook

Handle Heteroskedasticity: WLS approach

The statistical inference method we developed so far is based upon the following assumption:

Assumptions For Getting the $\hat{\beta}_{ols}$ as BLUE

4. Spherical Error Variance: $\mathrm{Var}[\boldsymbol{u}|\boldsymbol{X}] = \sigma^2 \boldsymbol{I}_n$
   - Homoskedasticity: $\mathrm{E}[u_i^2|\boldsymbol{X}] = \sigma^2 \quad \forall\, i = 1, 2, \cdots, n$
   - No Serial Correlation: $\mathrm{E}[u_i u_j|\boldsymbol{X}] = 0 \quad \forall\, i, j = 1, 2, \cdots, n;\ i \neq j$

$$\mathrm{Var}[\boldsymbol{u}|\boldsymbol{X}] = \begin{bmatrix} \sigma^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

Relexing the Homoskedasticity assumption to the Heteroskedasiticity assumption while holding the no serial correlation assumption same.

Assumptions For Getting the $\hat{\beta}_{wls}$ as BLUE

4. Ecliptic Variance: $\mathrm{Var}[u_i|\boldsymbol{x}_i] = \boldsymbol{\Lambda}$
   - Heteroskedasticity: $\mathrm{E}[u_i^2|\boldsymbol{x}_i] = v(\boldsymbol{x}_i) \quad \forall\, i = 1, 2, \cdots, n$
   - No Serial Correlation: $\mathrm{E}[u_i u_j|\boldsymbol{X}] = 0 \quad \forall\, i, j = 1, 2, \cdots, n;\ i \neq j$

$$\mathrm{Var}[\boldsymbol{u}|\boldsymbol{X}] = \boldsymbol{\Lambda} = \begin{bmatrix} v(\boldsymbol{x}_1) & 0 & 0 & \dots & 0 \\ 0 & v(\boldsymbol{x}_2) & 0 & \dots & 0 \\ 0 & 0 & v(\boldsymbol{x}_3) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & v(\boldsymbol{x}_n) \end{bmatrix}$$

The $v(\cdot)$ here is denoted as a functional form of the variance of the uncertainty u, where the input is the $i-$th observational sample $\boldsymbol{x}_i$

Heteroskedasiticity with Known Functional Form

Suppose the functional form $v(\cdot)$ is known, then given the sample data $\boldsymbol{x}_i$, the conditional variance is known and denoted as v_i where
$v_i = \mathrm{Var}[u_i|\boldsymbol{x}_i] = v(\boldsymbol{x}_i)$

$$\mathrm{Var}[\boldsymbol{u}|\boldsymbol{X}] = \boldsymbol{\Lambda} = \begin{bmatrix} v_1 & 0 & 0 & \dots & 0 \\ 0 & v_2 & 0 & \dots & 0 \\ 0 & 0 & v_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & v_n \end{bmatrix}$$

Since $\boldsymbol{\Lambda}$ is covariance matrix, it is nonsingular ($\boldsymbol{\Lambda}^{-1}$ exist), real and symmetric. By the eigen-decomposition we have $\boldsymbol{\Lambda}^{-1} = \boldsymbol{C}'\boldsymbol{C}$. Where $\boldsymbol{C}$ is nonsingular matrix.

Decompose Λ^{-1} the inverse Covariance Matrix for $\boldsymbol{u}$

For the above heteroskedasiticity case, it is easy to decompose Λ^{-1} to obtain the explicit expression for $\boldsymbol{C}$:

$$
\begin{aligned}
\Lambda^{-1} &= \begin{bmatrix} \frac{1}{v_1} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{v_2} & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{v_3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{v_n} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\sqrt{v_1}} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{v_2}} & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{\sqrt{v_3}} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{\sqrt{v_n}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{v_1}} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{v_2}} & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{\sqrt{v_3}} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{\sqrt{v_n}} \end{bmatrix} \\
&= \boldsymbol{C}'\boldsymbol{C}
\end{aligned}
$$

Transform the the Original Model

Motivation

Originally, the linear model of interest is $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$ or equivalently
$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{ip} x_p + u_i \quad i = 1, 2, \cdots n$
Since homoskedasticity in assumption 4 not hold anymore, the OLS estimator we previously used, $\hat{\boldsymbol{\beta}}_{ols} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$, is not BLUE. But if assumption 1-3 holds, the $\hat{\boldsymbol{\beta}}_{ols}$ is still unbiased and consistant. In order to get the BLUE estimator, we need to transform the original linear model by modified the sample data so that the transformed model still satisfying homoskedasticity.

Model Transformation and BLUE Estimator $\hat{\boldsymbol{\beta}}_{wls}$

Let $\tilde{\boldsymbol{Y}} = \boldsymbol{C}\boldsymbol{Y}, \tilde{\boldsymbol{X}} = \boldsymbol{C}\boldsymbol{X}, \tilde{\boldsymbol{u}} = \boldsymbol{C}\boldsymbol{u}$, then $\tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{X}}\beta + \tilde{\boldsymbol{u}}$
Since $\boldsymbol{C}$ is known, we claim the new model satisfy assumption 1-4.
Applying the OLS estimation procedure to the modified data we have the BLUE estimator: $\hat{\boldsymbol{\beta}}_{wls} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}}$

Compute the $\hat{\beta}_{wls}$ and its Conditional Variance

For the theoretical illustration purpose we decompose the inverse of the covariance matrix $\mathbf{\Lambda}^{-1} = \boldsymbol{C}'\boldsymbol{C}$, and modified the data. In the real computation of $\hat{\beta}_{wls}$, decomposition is not necessary.

$$
\begin{aligned}
\hat{\beta}_{wls} &= (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} \\
&= \left((\boldsymbol{C}\boldsymbol{X})'\boldsymbol{C}\boldsymbol{X}\right)^{-1}(\boldsymbol{C}\boldsymbol{X})'\boldsymbol{C}\boldsymbol{Y} \\
&= \left(\boldsymbol{X}'\boldsymbol{C}'\boldsymbol{C}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{C}'\boldsymbol{C}\boldsymbol{Y} \\
&= \left(\boldsymbol{X}'\Lambda^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\Lambda^{-1}\boldsymbol{Y} \\
\mathrm{Var}[\,\hat{\beta}_{wls}\,|\,\boldsymbol{X}\,] &= \left(\boldsymbol{X}'\Lambda^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\Lambda^{-1}\mathrm{Var}[\,\boldsymbol{Y}\,|\,\boldsymbol{X}\,]\Lambda^{-1}\boldsymbol{X}\left(\boldsymbol{X}'\Lambda^{-1}\boldsymbol{X}\right)^{-1} \\
&= \left(\boldsymbol{X}'\Lambda^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\Lambda^{-1}\Lambda\Lambda^{-1}\boldsymbol{X}\left(\boldsymbol{X}'\Lambda^{-1}\boldsymbol{X}\right)^{-1} \\
&= \left(\boldsymbol{X}'\Lambda^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\Lambda^{-1}\boldsymbol{X}\left(\boldsymbol{X}'\Lambda^{-1}\boldsymbol{X}\right)^{-1} \\
&= \left(\boldsymbol{X}'\Lambda^{-1}\boldsymbol{X}\right)^{-1}
\end{aligned}
$$

Heteroskedasiticity with Unknown Functional Form

Suppose the functional form $v(\cdot)$ is unknown, we can approximate $v(\cdot)$ by a linear relationship $v(\boldsymbol{x}_i) = \boldsymbol{\alpha}\boldsymbol{x}_i + \eta$. The linear relationship parameter α could be estimated by OLS regression on the estimated $\hat{u}_i^2$ and $\boldsymbol{x}_i$. Let $\hat{\boldsymbol{\alpha}}$ be the estimated $\boldsymbol{\alpha}$ then $\hat{v}_i = \hat{\boldsymbol{\alpha}}\boldsymbol{x}_i$. So the estimated covariance matrix is

$$\hat{\mathbf{\Lambda}} = \begin{bmatrix} \hat{v}_1 & 0 & 0 & \dots & 0 \\ 0 & \hat{v}_2 & 0 & \dots & 0 \\ 0 & 0 & \hat{v}_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \hat{v}_n \end{bmatrix}$$

When calculate the $\hat{\boldsymbol{\beta}}_{wls}$ and $\mathrm{Var}[\,\hat{\boldsymbol{\beta}}_{wls} \,|\, \boldsymbol{X}\,]$, the unknown covariance matrix $\mathrm{Var}[\boldsymbol{u}|\boldsymbol{X}] = \mathbf{\Lambda}$ could be replaced by the estimated value $\hat{\mathbf{\Lambda}}$.

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{wls} &= \left(\boldsymbol{X}'\hat{\mathbf{\Lambda}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\hat{\mathbf{\Lambda}}^{-1}\boldsymbol{Y} \\ \mathrm{Var}[\,\hat{\boldsymbol{\beta}}_{wls} \,|\, \boldsymbol{X}\,] &= \left(\boldsymbol{X}'\hat{\mathbf{\Lambda}}^{-1}\boldsymbol{X}\right)^{-1} \end{aligned}$$

Algorithm of WLS Estimation for Unknown Functional Form

- Conduct the regular OLS estimation to obtain $\hat{\beta}_{ols}$ and residuals $\hat{u}_i$
- Regress the sqared residuals $\hat{u}_i^2$ on $\boldsymbol{x}_i$ to obtain linear relationship coefficient $\hat{\alpha}$
- Calculate each estimated conditional variance $\hat{v}_i = \hat{\alpha}\boldsymbol{x}_i$ and form the estimated covariance matrix $\hat{\boldsymbol{\Lambda}}$
- Compute the WLS estimator $\hat{\beta}_{wls} = \left(\boldsymbol{X}'\hat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\hat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{Y}$

Handle Heteroskedasticity: GLS approach

If we relex the assumption further by allowing serial correlation, then we have the most general case of covariance matrix for the uncertainty u.

Most General Case for the Covariance Matrix for $\boldsymbol{u}$

4. General Covariance Matrix Form: $\mathrm{Var}[u_i|\boldsymbol{x}_i] = \boldsymbol{\Lambda}$
   - Heteroskedasticity: $\mathrm{E}[u_i^2|\boldsymbol{x}_i] = v(\boldsymbol{x}_i) \quad \forall\, i = 1, 2, \cdots, n$
   - Serial Correlation: $\mathrm{E}[u_i u_j|\boldsymbol{X}] = v(\boldsymbol{x}_i, \boldsymbol{x}_j) \quad \forall\, i, j = 1, 2, \cdots, n;\ i \neq j$

$$\mathrm{Var}[\boldsymbol{u}|\boldsymbol{X}] = \boldsymbol{\Lambda} = \begin{bmatrix} v(\boldsymbol{x}_1) & v(\boldsymbol{x}_1, \boldsymbol{x}_2) & v(\boldsymbol{x}_1, \boldsymbol{x}_3) & \dots & v(\boldsymbol{x}_1, \boldsymbol{x}_n) \\ v(\boldsymbol{x}_2, \boldsymbol{x}_1) & v(\boldsymbol{x}_2) & v(\boldsymbol{x}_2, \boldsymbol{x}_3) & \dots & v(\boldsymbol{x}_2, \boldsymbol{x}_n) \\ v(\boldsymbol{x}_3, \boldsymbol{x}_1) & v(\boldsymbol{x}_3, \boldsymbol{x}_2) & v(\boldsymbol{x}_3) & \dots & v(\boldsymbol{x}_3, \boldsymbol{x}_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v(\boldsymbol{x}_n, \boldsymbol{x}_1) & v(\boldsymbol{x}_n, \boldsymbol{x}_2) & v(\boldsymbol{x}_n, \boldsymbol{x}_3) & \dots & v(\boldsymbol{x}_n) \end{bmatrix}$$

The $v(\cdot)$ is the functional form of the variance and covariance of the uncertainty u. The input is the $i-$th and $j-$th observational sample $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$

Heteroskedasiticity and Serial Correlation with Known Functional Form

Suppose the functional form $v(\cdot)$ is known, then given the sample data $\boldsymbol{x}_i$, the conditional variance and covariance is known and denoted as v_{ii} and v_{ij} where $v_{ii} = \mathrm{Var}[u_i|\boldsymbol{x}_i] = v(\boldsymbol{x}_i)$ and $v_{ij} = \mathrm{Cov}[u_i, u_j|\boldsymbol{x}_i] = v(\boldsymbol{x}_i, \boldsymbol{x}_j)$

$$\mathrm{Var}[\boldsymbol{u}|\boldsymbol{X}] = \boldsymbol{\Lambda} = \begin{bmatrix} v_{11} & v_{12} & v_{13} & \dots & v_{1n} \\ v_{21} & v_{22} & v_{23} & \dots & v_{2n} \\ v_{31} & v_{32} & v_{33} & \dots & v_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & v_{n3} & \dots & v_{nn} \end{bmatrix}$$

Since $\boldsymbol{\Lambda}$ is covariance matrix, it is nonsingular ($\boldsymbol{\Lambda}^{-1}$ exist), real and symmetric. By the eigen-decomposition we have $\boldsymbol{\Lambda}^{-1} = \boldsymbol{C}'\boldsymbol{C}$. Where $\boldsymbol{C}$ is nonsingular matrix.

Transform the the Original Model

Motivation

Originally, the linear model of interest is $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{u}$ or equivalently $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{ip} x_p + u_i \quad i = 1, 2, \cdots n$
Since homoskedasticity in assumption 4 not hold anymore, the OLS estimator we previously used, $\hat{\boldsymbol{\beta}}_{ols} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$, is not BLUE. But if assumption 1-3 holds, the $\hat{\boldsymbol{\beta}}_{ols}$ is still unbiased and consistant. In order to get the BLUE estimator, we need to transform the original linear model by modified the sample data so that the transformed model still satisfying homoskedasticity.

Model Transformation and BLUE Estimator $\hat{\boldsymbol{\beta}}_{gls}$

Let $\tilde{\boldsymbol{Y}} = \boldsymbol{C}\boldsymbol{Y}, \tilde{\boldsymbol{X}} = \boldsymbol{C}\boldsymbol{X}, \tilde{\boldsymbol{u}} = \boldsymbol{C}\boldsymbol{u}$, then $\tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{X}}\beta + \tilde{\boldsymbol{u}}$
Since $\boldsymbol{C}$ is known, we claim the new model satisfy assumption 1-4.
Applying the OLS estimation procedure to the modified data we have the BLUE estimator: $\hat{\boldsymbol{\beta}}_{gls} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}}$

Compute the $\hat{\beta}_{gls}$ and its Conditional Variance

For the theoretical illustration purpose we decompose the inverse of the covariance matrix $\Lambda^{-1} = \boldsymbol{C}'\boldsymbol{C}$, and modified the data. In the real computation of $\hat{\beta}_{gls}$, decomposition is not necessary.

$$
\begin{aligned}
\hat{\beta}_{gls} &= (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} \\
&= \left((\boldsymbol{CX})'\boldsymbol{CX}\right)^{-1}(\boldsymbol{CX})'\boldsymbol{CY} \\
&= \left(\boldsymbol{X}'\boldsymbol{C}'\boldsymbol{CX}\right)^{-1}\boldsymbol{X}'\boldsymbol{C}'\boldsymbol{CY} \\
&= \left(\boldsymbol{X}'\Lambda^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\Lambda^{-1}\boldsymbol{Y} \\
\mathrm{Var}[\,\hat{\beta}_{gls}\,|\,\boldsymbol{X}\,] &= \left(\boldsymbol{X}'\Lambda^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\Lambda^{-1}\mathrm{Var}[\,\boldsymbol{Y}\,|\,\boldsymbol{X}\,]\Lambda^{-1}\boldsymbol{X}\left(\boldsymbol{X}'\Lambda^{-1}\boldsymbol{X}\right)^{-1} \\
&= \left(\boldsymbol{X}'\Lambda^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\Lambda^{-1}\Lambda\Lambda^{-1}\boldsymbol{X}\left(\boldsymbol{X}'\Lambda^{-1}\boldsymbol{X}\right)^{-1} \\
&= \left(\boldsymbol{X}'\Lambda^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\Lambda^{-1}\boldsymbol{X}\left(\boldsymbol{X}'\Lambda^{-1}\boldsymbol{X}\right)^{-1} \\
&= \left(\boldsymbol{X}'\Lambda^{-1}\boldsymbol{X}\right)^{-1}
\end{aligned}
$$

Handle Heteroskedasticity: Robust Standard Error

Homoskedasticity Conditional Variance for $\hat{\beta}_{ols}$

$$
\begin{aligned}
\hat{\beta}_{ols} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \\
\mathrm{Var}\left[\hat{\boldsymbol{\beta}}_{ols} \mid \boldsymbol{X}\right] &= \mathrm{E}\left[\left(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta}\right)' \mid \boldsymbol{X}\right] \\
&= \mathrm{E}\left[\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\right)\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\right)' \mid \boldsymbol{X}\right] \\
&= \mathrm{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\boldsymbol{u}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \mid \boldsymbol{X}\right] \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathrm{E}\left[\boldsymbol{u}\boldsymbol{u}' \mid \boldsymbol{X}\right]\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathrm{Var}\left[\boldsymbol{u} \mid \boldsymbol{X}\right]\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\sigma^2\boldsymbol{I}_n\,\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}
\end{aligned}
$$

If n is very large, relexing the Homoskedasticity assumption to the Heteroskedasiticity assumption while holding the no serial correlation assumption true.

Assumptions For Getting the $\hat{\beta}_{wls}$ as BLUE

4. Ecliptic Variance: $\mathrm{Var}[u_i|\boldsymbol{x}_i] = \boldsymbol{\Lambda}$
   - Heteroskedasticity: $\mathrm{E}[u_i^2|\boldsymbol{x}_i] = v(\boldsymbol{x}_i) \quad \forall\, i = 1, 2, \cdots, n$
   - No Serial Correlation: $\mathrm{E}[u_i u_j|\boldsymbol{X}] = 0 \quad \forall\, i, j = 1, 2, \cdots, n;\ i \neq j$

$$\mathrm{Var}[\boldsymbol{u}|\boldsymbol{X}] = \boldsymbol{\Lambda} = \begin{bmatrix} v(\boldsymbol{x}_1) & 0 & 0 & \dots & 0 \\ 0 & v(\boldsymbol{x}_2) & 0 & \dots & 0 \\ 0 & 0 & v(\boldsymbol{x}_3) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & v(\boldsymbol{x}_n) \end{bmatrix}$$

The $v(\cdot)$ here is denoted as a functional form of the variance of the uncertainty u, where the input is the $i-$th observational sample $\boldsymbol{x}_i$

Heteroskedasticity Robust Conditional Variance for $\hat{\beta}_{ols}$ with Known Functional Form

If the functional form $v(\cdot)$ is known, then $\mathrm{Var}[\boldsymbol{u}|\boldsymbol{X}] = \boldsymbol{\Lambda}$ is known. The robust conditional variance could be computed directly by following the definition of covariance matrix.

$$
\begin{aligned}
\hat{\beta}_{ols} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \\
\mathrm{Var}\left[\hat{\boldsymbol{\beta}}_{ols} \,|\, \boldsymbol{X}\right] &= \mathrm{E}\left[\left(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta}\right)' \,|\, \boldsymbol{X}\right] \\
&= \mathrm{E}\left[\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\right)\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\right)' \,|\, \boldsymbol{X}\right] \\
&= \mathrm{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\boldsymbol{u}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \,|\, \boldsymbol{X}\right] \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathrm{E}\left[\boldsymbol{u}\boldsymbol{u}' \,|\, \boldsymbol{X}\right]\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathrm{Var}\left[\boldsymbol{u} \,|\, \boldsymbol{X}\right]\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\Lambda\,\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}
\end{aligned}
$$

By CLT: $\hat{\boldsymbol{\beta}}_{ols} \sim N\left(\boldsymbol{\beta},\ (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\Lambda\,\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right)$

$$\hat{\boldsymbol{\beta}}_{ols} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_p \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \begin{bmatrix} m_{(00)} & m_{(01)} & m_{(02)} & \cdots & m_{(0p)} \\ m_{(10)} & m_{(11)} & m_{(12)} & \cdots & m_{(1p)} \\ m_{(20)} & m_{(21)} & m_{(22)} & \cdots & m_{(2p)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{(p0)} & m_{p1)} & m_{(p2)} & \cdots & m_{(pp)} \end{bmatrix} \right)$$

where

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\Lambda\,\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{bmatrix} m_{(00)} & m_{(01)} & m_{(02)} & \cdots & m_{(0p)} \\ m_{(10)} & m_{(11)} & m_{(12)} & \cdots & m_{(1p)} \\ m_{(20)} & m_{(21)} & m_{(22)} & \cdots & m_{(2p)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{(p0)} & m_{p1)} & m_{(p2)} & \cdots & m_{(pp)} \end{bmatrix}$$

The robusted standard deviation of the k-th entry of $\hat{\boldsymbol{\beta}}_{ols}$ is $\sqrt{m_{(kk)}}$

Heteroskedasticity Robust Conditional Variance for $\hat{\beta}_{ols}$ with Unknown Functional Form

If the functional form $v(\cdot)$ is Unknown, then $\mathrm{Var}[\boldsymbol{u}|\boldsymbol{X}] = \boldsymbol{\Lambda}$ could be estimated by the diagonal matrix $\hat{\boldsymbol{\Lambda}}$, where each entry on the diagonal line is the corresponding sample squared residuals $\hat{u}_i^2$ as follow:

$$\hat{\boldsymbol{\Lambda}} = \begin{bmatrix} \hat{u}_1^2 & 0 & 0 & \dots & 0 \\ 0 & \hat{u}_2^2 & 0 & \dots & 0 \\ 0 & 0 & \hat{u}_3^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \hat{u}_n^2 \end{bmatrix}$$

Assumptions For Getting the consistent estimator for $\boldsymbol{X}'\,\boldsymbol{\Lambda}\,\boldsymbol{X}$

6. No outlier for each regressor x_j: $\mathrm{E}[x_j^4] < \infty \quad j = 1, 2, \cdots, p$

with the above assumption, we have $\boldsymbol{X}'\,\hat{\boldsymbol{\Lambda}}\,\boldsymbol{X} \xrightarrow{p} \boldsymbol{X}'\,\boldsymbol{\Lambda}\,\boldsymbol{X}$

Heteroskedasticity Robust Conditional Variance for $\hat{\beta}_{ols}$ with Unknown Functional Form

The robust conditional variance could be computed by following the definition of covariance matrix and replacing $\boldsymbol{X}'\mathrm{Var}[\boldsymbol{u}|\boldsymbol{X}]\,\boldsymbol{X} = \boldsymbol{X}'\,\Lambda\,\boldsymbol{X}$ by $\boldsymbol{X}'\,\hat{\Lambda}\,\boldsymbol{X}$

$$
\begin{aligned}
\hat{\beta}_{ols} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \\
\mathrm{Var}\left[\hat{\beta}_{ols}\,|\,\boldsymbol{X}\right] &= \mathrm{E}\left[\left(\hat{\beta}_{ols}-\boldsymbol{\beta}\right)\left(\hat{\beta}_{ols}-\boldsymbol{\beta}\right)'\,|\,\boldsymbol{X}\right] \\
&= \mathrm{E}\left[\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\right)\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\right)'\,|\,\boldsymbol{X}\right] \\
&= \mathrm{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\boldsymbol{u}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\,|\,\boldsymbol{X}\right] \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathrm{E}\left[\boldsymbol{u}\boldsymbol{u}'\,|\,\boldsymbol{X}\right]\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathrm{Var}\left[\boldsymbol{u}\,|\,\boldsymbol{X}\right]\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\hat{\Lambda}\,\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}
\end{aligned}
$$

By CLT: $\hat{\beta}_{ols} \sim N\Big(\beta,\ (X'X)^{-1}X'\,\hat{\Lambda}\,X(X'X)^{-1}\Big)$

$$\hat{\beta}_{ols} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_p \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \begin{bmatrix} \hat{m}_{(00)} & \hat{m}_{(01)} & \hat{m}_{(02)} & \cdots & \hat{m}_{(0p)} \\ \hat{m}_{(10)} & \hat{m}_{(11)} & \hat{m}_{(12)} & \cdots & \hat{m}_{(1p)} \\ \hat{m}_{(20)} & \hat{m}_{(21)} & \hat{m}_{(22)} & \cdots & \hat{m}_{(2p)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{m}_{(p0)} & \hat{m}_{p1)} & \hat{m}_{(p2)} & \cdots & \hat{m}_{(pp)} \end{bmatrix} \right)$$

where

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\hat{\Lambda}\,\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{bmatrix} \hat{m}_{(00)} & \hat{m}_{(01)} & \hat{m}_{(02)} & \cdots & \hat{m}_{(0p)} \\ \hat{m}_{(10)} & \hat{m}_{(11)} & \hat{m}_{(12)} & \cdots & \hat{m}_{(1p)} \\ \hat{m}_{(20)} & \hat{m}_{(21)} & \hat{m}_{(22)} & \cdots & \hat{m}_{(2p)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{m}_{(p0)} & \hat{m}_{p1)} & \hat{m}_{(p2)} & \cdots & \hat{m}_{(pp)} \end{bmatrix}$$

The robusted standard deviation of the k-th entry of $\hat{\beta}_{ols}$ is $\sqrt{\hat{m}_{(kk)}}$

Connect to the Assumptions in Textbook

Textbook Assumptions for Causal Inference

1. Underlying population space (X, Y) has a linear relationship $y = \boldsymbol{x}'\boldsymbol{\beta} + u = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + u$
2. No Multicollinearity
3. Strict Exogeneity Condition: $\mathrm{E}[\boldsymbol{u}|\boldsymbol{X}] = \boldsymbol{0}$
4. sample $(\boldsymbol{x}_i, y_i)$ are i.i.d
5. No outlier condition: $\mathrm{E}[\boldsymbol{x}^4] < \infty$ and $\mathrm{E}[y^4] < \infty$

Assumption 1-3 is same as what we proposed.

- From Assumption 1-2 we have $\hat{\boldsymbol{\beta}}_{ols}$ exists and $\hat{\boldsymbol{\beta}}_{ols} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$
- From Assumption 1-3 we have $\hat{\boldsymbol{\beta}}_{ols}$ unbiased: $\mathrm{E}[\hat{\boldsymbol{\beta}}_{ols}|\boldsymbol{X}] = \boldsymbol{\beta}$ and also consistent (if uncertainty $\boldsymbol{u}$ has a finite variance): $\hat{\boldsymbol{\beta}}_{ols} \xrightarrow{p} \boldsymbol{\beta}$

Assumption 4 act like randomized controled experiment.

- Gives the causal relationship between $\boldsymbol{x}$ and y
- No Serial Correlation : $\mathrm{E}[u_i u_j|\boldsymbol{X}] = 0 \quad \forall\, i, j = 1, 2, \cdots, n;\ i \neq j$

Assumption 5 and Large sample size n makes the statistical inference possible

- Assumption 4 gives the No serial correlation. Thus the underlying Ecliptic Variance Assumption $\mathrm{Var}[u_i|\boldsymbol{x}_i] = \boldsymbol{\Lambda}$ holds. Combined with $\mathrm{E}[\boldsymbol{x}^4] < \infty$ in Assumption 5, the covariance matrix of $\mathrm{Var}\left[\hat{\boldsymbol{\beta}}_{ols}\,|\,\boldsymbol{X}\right] = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\hat{\boldsymbol{\Lambda}}\,\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$ could be calculated. The original z statistics for the k-th entry could be constructed by subtract the true β_k from $\hat{\beta}_k$ and devide the robusted standard deviation: $z = \frac{\hat{\beta}_k - \beta_k}{\sqrt{\hat{m}_{(kk)}}}$
- $\mathrm{E}[\boldsymbol{x}^4] < \infty$ and $\mathrm{E}[y^4] < \infty$ implies $\mathrm{E}[u^4] < \infty$. Thus $\mathrm{E}[u^2] < \infty$ i.e u has a finite variance.