Non-spherical disturbance

When the variance-covariance matrix of the error terms is not spherical, we call the Classical linear regression (CLR) as General linear regression (GLR) model. e.g.,

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{u}, \boldsymbol{u} \sim \boldsymbol{IID}(0, \boldsymbol{\Omega})$$

where $\boldsymbol{\Omega}$ the variance-covariance matrix of the error term, sometimes it is written as an alternative notation, e.g., $\mathbf{G}$.

In this case, OLS is still unbiased.

However, the variance-covariance matrix for the parameter $\boldsymbol{\beta}^{\mathrm{OLS}}$ is incorrect, thus $Var(\boldsymbol{\beta}^{\mathrm{OLS}})$ is biased (usually downwards) even asymptotically. Therefore, interval estimation and hypothesis testing for the parameter are unreliable.

A mitigation is to use heteroskedasticiy-consistent and autocorrelation-consistent variance-covariance matrix estimators for the OLS estimator to eliminate the asymptotic bias (but not completely the small sample bias) of $Var(\boldsymbol{\beta}^{\mathrm{OLS}})$. This is sometimes called a robust variance estimator.

Efficiency: $\boldsymbol{\beta}^{\mathrm{GLS}}$ becomes the BLUE estimator. $\boldsymbol{\beta}^{\mathrm{GLS}}$ uses information more efficiently compared to $\boldsymbol{\beta}^{\mathrm{OLS}}$. e.g., we can use the information that some disturbances are likely to be large because their variance are large. Therefore, instead of minimizing the sum of squared residuals, we can minimize weighted sum of squared residuals. Observations whose residuals are expected to be large because other residuals are large are given smaller weights.

If we can assume the error terms are jointly normally distributed, $\boldsymbol{\beta}^{\mathrm{GLS}}$ replaces $\boldsymbol{\beta}^{\mathrm{OLS}}$ to become equivalent to the maximum likely estimator.

However, we need to know $\boldsymbol{\Omega}$ to calculate $\boldsymbol{\beta}^{\mathrm{GLS}}$. As $\boldsymbol{\Omega}$ is rarely known, it is tempting to simply employ $\boldsymbol{\beta}^{\mathrm{OLS}}$ with heteroskedasticiy-consistent and autocorrelation-consistent variance-covariance matrix.

Thus, we tend to employ $\boldsymbol{\beta}^{\mathrm{GLS}}$ when we somehow know a little bit about $\boldsymbol{\Omega}$ and / or the non-spherical disturbance term cause substantive consequence. What we do is that we invent the **Feasible GLS** (e.g., **FGLS**) estimator by consistently estimating $\boldsymbol{\Omega}$. Thus, we somehow take into account the nonsphericalness of the disturbance. As a result, with FGLS, $Var(\boldsymbol{\beta}^{\mathrm{FGLS}})$is asymptotically unbiased. However, note that FGLS is biased and nonlinear.