Violation of the homoskedasticity assumption, the GLM, the GLS estimator, HCV estimator, and the tests of heteroskedasticity

Empirical Methods, Lecture 6

Tunga Kantarcı, FEB, Groningen University, Spring 2025

SLM, spherical errors assumption

Consider the regression model,

$$y_i = x_i\beta + \varepsilon_i.$$

In the SLM, we posed assumption A4. That is, we assumed that ε_i is spherical. That is, it is homoskedastic and serially uncorrelated.

SLM, homoskedasticity assumption

$$\begin{aligned}
\mathrm{Var}\left[\varepsilon_i \mid \boldsymbol{X}\right] &= \mathrm{E}\left[\varepsilon_i \varepsilon_i \mid \boldsymbol{X}\right] - \mathrm{E}\left[\varepsilon_i \mid \boldsymbol{X}\right] \mathrm{E}\left[\varepsilon_i \mid \boldsymbol{X}\right] \\
&= \mathrm{E}\left[\varepsilon_i \varepsilon_i \mid \boldsymbol{X}\right] \\
&= \sigma^2
\end{aligned}$$

if $\mathrm{E}\left[\varepsilon_i \mid \boldsymbol{X}\right] = 0$. Homoskedasticity states that ε_i has the same variance σ^2 at all observations in $\boldsymbol{X}$.

SLM, nonautocorrelation assumption

$$\begin{aligned}\operatorname{Cov}[\varepsilon_i, \varepsilon_j \mid \boldsymbol{X}] &= \operatorname{E}[\varepsilon_i \varepsilon_j \mid \boldsymbol{X}] - \operatorname{E}[\varepsilon_i \mid \boldsymbol{X}] \operatorname{E}[\varepsilon_j \mid \boldsymbol{X}] \\ &= \operatorname{E}[\varepsilon_i \varepsilon_j \mid \boldsymbol{X}] \\ &= 0\end{aligned}$$

if $\operatorname{E}[\varepsilon_i \mid \boldsymbol{X}] = 0$. Nonautocorrelation states that ε_i is uncorrelated with every other ε_j at all observations in $\boldsymbol{X}$.

SLM, spherical errors assumption

For a given error, ε_i, the variance, conditional on $\boldsymbol{X}$, is

$$\mathrm{E}\left[\varepsilon_i \varepsilon_i \mid \boldsymbol{X}\right] = \sigma^2,$$

and the covariance, conditional on $\boldsymbol{X}$, is

$$\mathrm{E}\left[\varepsilon_i \varepsilon_j \mid \boldsymbol{X}\right] = 0.$$

SLM, spherical errors assumption

For n errors, $\boldsymbol{\varepsilon}$, the variance-covariance matrix is

$$
\begin{aligned}
\operatorname{Var}[\boldsymbol{\varepsilon} \mid \boldsymbol{X}] &= \operatorname{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}\right] - \operatorname{E}[\boldsymbol{\varepsilon} \mid \boldsymbol{X}] \operatorname{E}\left[\boldsymbol{\varepsilon}' \mid \boldsymbol{X}\right] \\
&= \operatorname{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}\right] \\
&= \sigma^2 I_n \\
&= \sigma^2 \boldsymbol{I}
\end{aligned}
$$

if $\operatorname{E}[\varepsilon_i \mid \boldsymbol{X}] = 0$.

$\boldsymbol{\varepsilon}$ is $n \times 1$. $\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'$ is $n \times n$. Hence, $\operatorname{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}]$ is $n \times n$. How does it look like?

SLM, spherical errors assumption

$$
\mathrm{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}\right] = \begin{bmatrix} \mathrm{E}\left[\varepsilon_1\varepsilon_1 \mid \boldsymbol{X}\right] & \mathrm{E}\left[\varepsilon_1\varepsilon_2 \mid \boldsymbol{X}\right] & \dots & \mathrm{E}\left[\varepsilon_1\varepsilon_n \mid \boldsymbol{X}\right] \\ \mathrm{E}\left[\varepsilon_2\varepsilon_1 \mid \boldsymbol{X}\right] & \mathrm{E}\left[\varepsilon_2\varepsilon_2 \mid \boldsymbol{X}\right] & \dots & \mathrm{E}\left[\varepsilon_2\varepsilon_n \mid \boldsymbol{X}\right] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}\left[\varepsilon_n\varepsilon_1 \mid \boldsymbol{X}\right] & \mathrm{E}\left[\varepsilon_n\varepsilon_2 \mid \boldsymbol{X}\right] & \dots & \mathrm{E}\left[\varepsilon_n\varepsilon_n \mid \boldsymbol{X}\right] \end{bmatrix}
$$

$$
= \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}
$$

$$
= \underbrace{\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}}_{I_n} \sigma^2
$$

GLM

What defines an econometric model is very much about the assumptions made about ε_i. If ε_i is spherical, the linear model is the standard linear model. If ε_i is nonspherical, the linear model is the generalised linear model. You now understand why we called it standard.

GLM

As we relax the spherical errors assumption, we can differentiate between two cases.

If we relax the assumption that ε_i is homoskedastic, then ε_i is heteroskedastic.

If we relax the assumption that ε_i is non-autocorrelated, then ε_i is autocorrelated.

GLM, error is heteroskedastic

We start with heteroskedasticity, and focus on it in this course.

GLM, error is heteroskedastic

For error i, ε_i,

$$\begin{aligned}
\operatorname{Var}[\varepsilon_i \mid \boldsymbol{X}] &= \operatorname{E}[\varepsilon_i \varepsilon_i \mid \boldsymbol{X}] - \operatorname{E}[\varepsilon_i \mid \boldsymbol{X}] \operatorname{E}[\varepsilon_i \mid \boldsymbol{X}] \\
&= \operatorname{E}[\varepsilon_i \varepsilon_i \mid \boldsymbol{X}] \\
&= \sigma_i^2 \\
&= \sigma^2 \omega_i
\end{aligned}$$

if $\operatorname{E}[\varepsilon_i \mid \boldsymbol{X}] = 0$.

ω_i is a function of x_i. Hence, the explicit notation is in fact $\omega(x_i)$. We use the former for ease of notation. $\sigma^2 \omega(x_i)$ says that the variance of ε_i depends on the different values of an explanatory variable in some given functional form. Mind the conditioning. We think of this as the error being drawn from a different distribution for each observation i of the explanatory variable.

GLM, error is heteroskedastic

For n errors, ε, the variance-covariance matrix is

$$\begin{aligned}
\operatorname{Var}[\boldsymbol{\varepsilon} \mid \boldsymbol{X}] &= \operatorname{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}\right] - \operatorname{E}[\boldsymbol{\varepsilon} \mid \boldsymbol{X}] \operatorname{E}\left[\boldsymbol{\varepsilon}' \mid \boldsymbol{X}\right] \\
&= \operatorname{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}\right] \\
&= \sigma^2 \boldsymbol{\Omega}
\end{aligned}$$

if $\operatorname{E}[\varepsilon_i \mid \boldsymbol{X}] = 0$.

$\boldsymbol{\Omega}$ is $n \times n$. It is a function of $\boldsymbol{X}$. Hence the explicit notation is in fact $\boldsymbol{\Omega}(\boldsymbol{X})$.

GLM, error is heteroskedastic

How does $\mathrm{E}\left[\varepsilon\varepsilon' \mid \boldsymbol{X}\right]$ look like?

GLM, error is heteroskedastic

$$\mathrm{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}\right] = \begin{bmatrix} \mathrm{E}\left[\varepsilon_1\varepsilon_1 \mid \boldsymbol{X}\right] & \mathrm{E}\left[\varepsilon_1\varepsilon_2 \mid \boldsymbol{X}\right] & \dots & \mathrm{E}\left[\varepsilon_1\varepsilon_n \mid \boldsymbol{X}\right] \\ \mathrm{E}\left[\varepsilon_2\varepsilon_1 \mid \boldsymbol{X}\right] & \mathrm{E}\left[\varepsilon_2\varepsilon_2 \mid \boldsymbol{X}\right] & \dots & \mathrm{E}\left[\varepsilon_2\varepsilon_n \mid \boldsymbol{X}\right] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}\left[\varepsilon_n\varepsilon_1 \mid \boldsymbol{X}\right] & \mathrm{E}\left[\varepsilon_n\varepsilon_2 \mid \boldsymbol{X}\right] & \dots & \mathrm{E}\left[\varepsilon_n\varepsilon_n \mid \boldsymbol{X}\right] \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \omega_n \end{bmatrix}}_{\boldsymbol{\Omega}} \sigma^2.$$

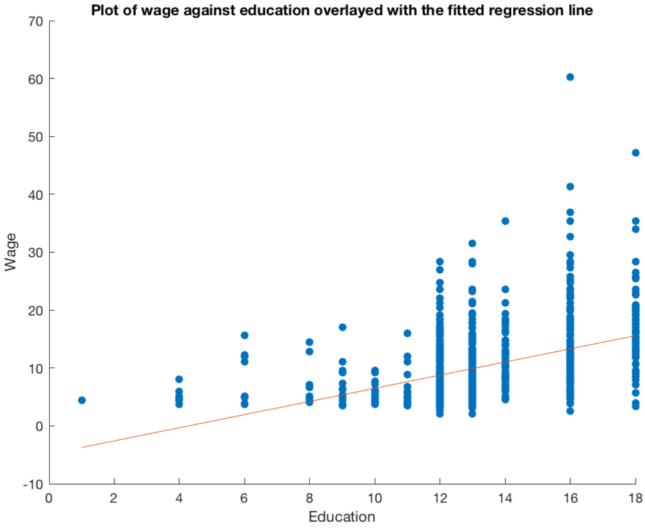
GLM, error is heteroskedastic

In Greek, hetero means different, and skedasis means dispersion. Different dispersion. Non-constant variance!

GLM, error is heteroskedastic, example

We want to explain wage with education.

GLM, error is heteroskedastic, example



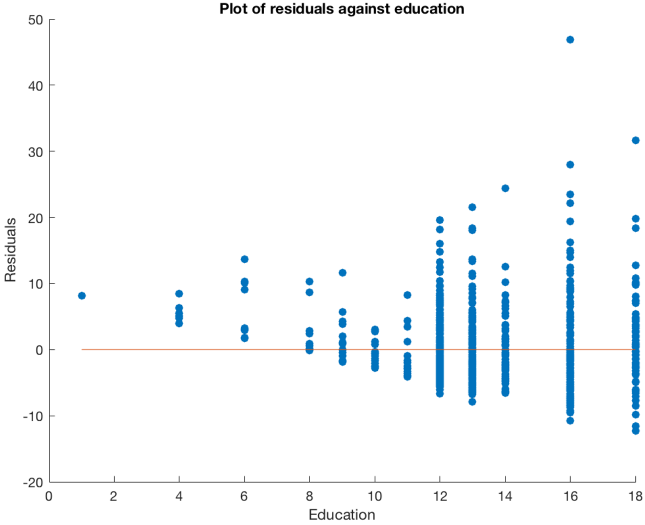
GLM, error is heteroskedastic, example

Why wage does not have a constant variance at given values of education? Think of the job opportunities. Probably more education means a wider variety of job opportunities. Then, wage is more variable at higher levels of education.

GLM, error is heteroskedastic, example

But it is difficult to observe the job opportunities people have. Hence, it enters the error. But if it enters the error, then, the errors will show more variation at higher levels of education. Hence, errors do not have a constant variance across given values of education. Variance of ε conditional on $\boldsymbol{X}$ is not constant!

GLM, error is heteroskedastic, example



GLM, error is heteroskedastic

Note that the example is based on sample data. By looking at the sample distribution of wage, or the residuals, against education, we try to infer whether heteroskedasticity is in play. Let's revert back to the population model.

GLM, error is heteroskedastic

If we suspect that the variance of wage, at given values of education is not constant, the variance of the error, at given values of education will not be constant. Why?

GLM, error is heteroskedastic

Consider the linear model

$$y_i = x_i\beta + \varepsilon_i.$$

Taking the expectation, conditional on x_i,

$$\mathrm{E}\left[y_i \mid x_i\right] = x_i\beta.$$

Rewriting the linear model,

$$y_i = \mathrm{E}\left[y_i \mid x_i\right] + \varepsilon_i.$$

The error represents dispersion around the conditional expectation function. Is this dispersion constant?

GLM, error is heteroskedastic

Dispersion is about variance. Then check the variance. Consider again the linear model

$$y_i = x_i\beta + \varepsilon_i.$$

Taking the variance, conditional on x_i,

$$\begin{aligned}
\color{red}{\mathrm{Var}\left[y_i \mid x_i\right]} &= \mathrm{Var}\left[x_i\beta \mid x_i\right] + \mathrm{Var}\left[\varepsilon_i \mid x_i\right] \\
&= \beta^2 \mathrm{Var}\left[x_i \mid x_i\right] + \mathrm{Var}\left[\varepsilon_i \mid x_i\right] \\
&\overset{!}{=} \color{red}{\mathrm{Var}\left[\varepsilon_i \mid x_i\right]} \\
&= \sigma_i^2.
\end{aligned}$$

GLM, error is autocorrelated

We continue with autocorrelation. The formal definition is

$$
\begin{aligned}
\operatorname{Cov}[\varepsilon_i, \varepsilon_j \mid \boldsymbol{X}] &= \operatorname{E}[\varepsilon_i \varepsilon_j \mid \boldsymbol{X}] - \operatorname{E}[\varepsilon_i \mid \boldsymbol{X}] \operatorname{E}[\varepsilon_j \mid \boldsymbol{X}] \\
&= \operatorname{E}[\varepsilon_i \varepsilon_j \mid \boldsymbol{X}] \neq 0
\end{aligned}
$$

if $\operatorname{E}[\varepsilon_i \mid \boldsymbol{X}] = 0$. This says that one unobserved factor is correlated with another. Then, ε_i is to be autocorrelated.

Skip.

GLM, error is autocorrelated

How does $\mathrm{E}\left[\varepsilon\varepsilon' \mid \boldsymbol{X}\right]$ look like?

Skip.

GLM, error is autocorrelated

$$
\mathrm{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}\right] = \begin{bmatrix} \mathrm{E}\left[\varepsilon_1\varepsilon_1 \mid \boldsymbol{X}\right] & \mathrm{E}\left[\varepsilon_1\varepsilon_2 \mid \boldsymbol{X}\right] & \dots & \mathrm{E}\left[\varepsilon_1\varepsilon_n \mid \boldsymbol{X}\right] \\ \mathrm{E}\left[\varepsilon_2\varepsilon_1 \mid \boldsymbol{X}\right] & \mathrm{E}\left[\varepsilon_2\varepsilon_2 \mid \boldsymbol{X}\right] & \dots & \mathrm{E}\left[\varepsilon_2\varepsilon_n \mid \boldsymbol{X}\right] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}\left[\varepsilon_n\varepsilon_1 \mid \boldsymbol{X}\right] & \mathrm{E}\left[\varepsilon_n\varepsilon_2 \mid \boldsymbol{X}\right] & \dots & \mathrm{E}\left[\varepsilon_n\varepsilon_n \mid \boldsymbol{X}\right] \end{bmatrix}
$$

$$
= \underbrace{\begin{bmatrix} 1 & \rho & \dots & \rho^{T-1} \\ \rho & 1 & \dots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \dots & 1 \end{bmatrix}}_{\boldsymbol{\Omega}} \sigma^2.
$$

Skip.

GLM, error is autocorrelated

$\mathbf{\Omega}$ results from the following example. Consider the linear model

$$y_t = x_t \beta + \varepsilon_t$$

where observations are realisations from different time periods. If

$$\varepsilon_t = \varepsilon_{t-1}\rho + \upsilon_t,$$

where $\upsilon_t \sim IID\left(0, \sigma_\upsilon^2\right)$, and $|\rho| < 1$, it can be shown that

$$\mathrm{E}\left[\varepsilon_t \varepsilon_t\right] = \sigma_\upsilon^2 / \left(1 - \rho^2\right) \equiv \sigma^2$$

and

$$\mathrm{E}\left[\varepsilon_t \varepsilon_s\right] = \sigma_\upsilon^2 / \left(1 - \rho^2\right) \rho^{|t-s|} = \rho^{|t-s|} \sigma^2.$$

Skip.

GLM

In this lecture we only consider the case of heteroskedasticity.

GLM, Model assumptions

A1. Linearity: the model is linear in $\boldsymbol{\beta}$.

A2. Full column rank: $rank\left(\boldsymbol{X}\right) = K$.

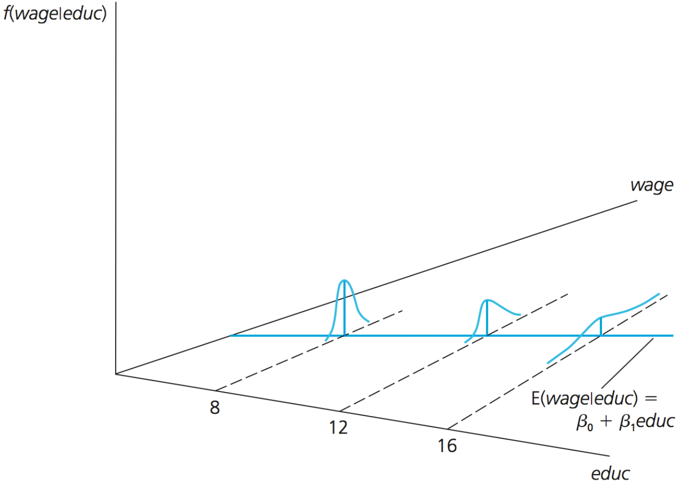
A3: Strict exogeneity: $\mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_k\right] = 0$. Hence, the conditional expectation function follows.

A4: Heteroskedasticity: $\mathrm{Var}\left[\varepsilon_i \mid \boldsymbol{X}\right] = \sigma_i^2$.

A5: The data $\{(\boldsymbol{x}_i, y_i) : i = 1, 2, \ldots, n\}$ is a random sample.

A6: We will assume that errors are normal if n is finite.

GLM, Model assumptions



GLM, OLS estimator

We use the OLS estimator to estimate the parameters of the SLM. Can we use the OLS estimator to estimate the parameters of the GLM where the error is heteroskedastic? Does $\hat{\beta}$ still have the desirable statistical properties?

GLM, OLS estimator

In this lecture $\hat{\boldsymbol{\beta}}_{OLS} \equiv \hat{\boldsymbol{\beta}}$.

GLM, OLS estimator is unbiased

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon.$$

Taking the expectation, conditional on $\boldsymbol{X}$,

$$\begin{aligned} \mathrm{E}\left[\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right] &= \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathrm{E}\left[\varepsilon \mid \boldsymbol{X}\right] \\ &= \boldsymbol{\beta} \end{aligned}$$

if $\mathrm{E}\left[\varepsilon \mid \boldsymbol{X}\right] = 0$.

This means that $\hat{\boldsymbol{\beta}}$ is still unbiased.

GLM, OLS estimator is not efficient

Taking the variance, conditional on $\boldsymbol{X}$,

$$\begin{aligned}
\operatorname{Var}\left[\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right] &= \operatorname{E}\left[\left(\hat{\boldsymbol{\beta}} - \operatorname{E}\left[\hat{\boldsymbol{\beta}}\right]\right)\left(\hat{\boldsymbol{\beta}} - \operatorname{E}\left[\hat{\boldsymbol{\beta}}\right]\right)' \,\middle|\, \boldsymbol{X}\right] \\
&= \operatorname{E}\left[\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)' \,\middle|\, \boldsymbol{X}\right] \\
&= \operatorname{E}\left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \,\middle|\, \boldsymbol{X}\right] \\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\operatorname{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}\right]\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\sigma^2\boldsymbol{\Omega}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}
\end{aligned}$$

since $\operatorname{E}\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta}$ by the LIE, and $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$.

GLM, OLS estimator is not efficient

Recall that in the SLM,

$$\mathrm{Var}\left[\boldsymbol{\varepsilon} \mid \boldsymbol{X}\right] = \sigma^2 \boldsymbol{I},$$

and

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

We have shown that

$$\mathrm{Var}\left[\hat{\boldsymbol{\beta}}_0 \mid \boldsymbol{X}\right] \geq \mathrm{Var}\left[\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right],$$

where $\hat{\boldsymbol{\beta}}_0$ is a competitor (linear, unbiased estimator) LUE. $\hat{\boldsymbol{\beta}}$ was BLUE because it is the estimator with the smallest variance in the SLM.

GLM, OLS estimator is not efficient

Now in the GLM,

$$\mathrm{Var}\left[\boldsymbol{\varepsilon} \mid \boldsymbol{X}\right] = \sigma^2 \boldsymbol{\Omega},$$

and

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

Is

$$\mathrm{Var}\left[\hat{\boldsymbol{\beta}}_0 \mid \boldsymbol{X}\right] \geq \mathrm{Var}\left[\hat{\boldsymbol{\beta}}_{\boldsymbol{OLS}} \mid \boldsymbol{X}\right]$$

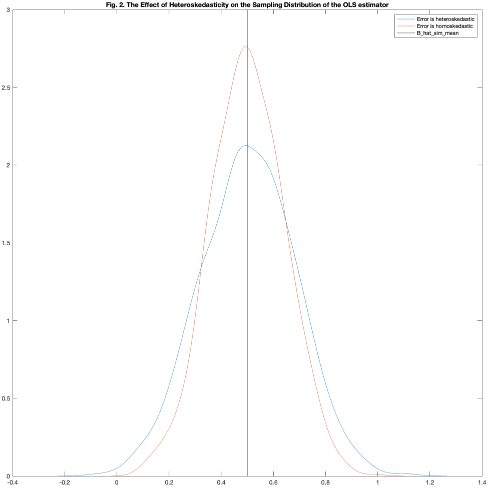
still true in the GLM? It is not! It can be shown that another estimator, $\hat{\boldsymbol{\beta}}_{GLS}$, is more efficient. That is,

$$\mathrm{Var}\left[\hat{\boldsymbol{\beta}}_0 \mid \boldsymbol{X}\right] \geq \mathrm{Var}\left[\hat{\boldsymbol{\beta}}_{GLS} \mid \boldsymbol{X}\right]$$

OLS estimator is not efficient in the GLM!

GLM, OLS estimator is not efficient

Sampling distribution of $\hat{\beta}_{OLS}$ when ε_i is heteroskedastic and when it is homoskedastic.



GLM, OLS estimator is not efficient, implications

As the figure shows, the sampling distribution of $\hat{\beta}$ has a larger variance in the GLM. What are the implications?

GLM, OLS estimator is not efficient, implications

$\hat{\beta}$ is not a precise estimator.

GLM, OLS estimator is not efficient, implications

In the SLM,

$$\mathrm{Var}\left[\boldsymbol{\varepsilon} \mid \boldsymbol{X}\right] = \sigma^2 \boldsymbol{I},$$

and

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

Using this variance, we derive and calculate the t and F statistics, and know that these statistics have the exact t and F distributions. That is, recall that

$$z_k = \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\sigma^2 S^{kk}}} \sim N\left[0, 1\right],$$

where

$$S^{kk} \equiv \left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{k,k}.$$

GLM, OLS estimator is not efficient, implications

Now in the GLM,

$$\text{Var}\left[\boldsymbol{\varepsilon} \mid \boldsymbol{X}\right] = \sigma^2 \boldsymbol{\Omega},$$

and

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

GLM, OLS estimator is not efficient, implications

When the error is heteroskedastic, we cannot use

$$\text{Var}\left[\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right] = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

to calculate the t and F statistics. This is not the correct variance to use for these statistics. Can or should we use

$$\text{Var}\left[\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right] = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

instead? Yes, we can, and should, if we use a consistent estimator of this variance.

Note that we need an estimator of this variance because σ is unobserved. We will derive this estimator later.

GLM, OLS estimator is normal

We know that

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}.$$

Assume that $\boldsymbol{\varepsilon}$ is multivariate normal. That is,

$$\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N\left[\mathbf{0}, \sigma^2\boldsymbol{\Omega}\right].$$

Is $\hat{\boldsymbol{\beta}}$ multivariate normal in the GLM?

GLM, OLS estimator is normal

We condition on $\boldsymbol{X}$ and hence treat it as given. The matrix

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}',$$

is $K \times n$. Recast it as a $K \times n$ matrix

$$\begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \dots & \boldsymbol{w}_n \end{bmatrix}.$$

$\boldsymbol{\varepsilon}$ is $n \times 1$. $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$ becomes

$$\boldsymbol{w}_1\varepsilon_1 + \boldsymbol{w}_2\varepsilon_2 + \ldots + \boldsymbol{w}_n\varepsilon_n.$$

Hence, $\hat{\boldsymbol{\beta}}$ is a linear combination of the elements of $\boldsymbol{\varepsilon}$. A linear combination of normal random variables is normal. Hence, $\hat{\boldsymbol{\beta}}$ is multivariate normal.

GLM, OLS estimator is normal

Using the mean and variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ derived above, we have

$$\hat{\boldsymbol{\beta}} \mid \boldsymbol{X} \sim N\left[\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right].$$

GLM, OLS estimator is consistent

In the GLM,

$$\mathsf{Var}\left[\boldsymbol{\varepsilon} \mid \boldsymbol{X}\right] = \sigma^2 \boldsymbol{\Omega},$$

and

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

GLM, OLS estimator is consistent

It can be shown that, in the GLM, the variance of $\hat{\boldsymbol{\beta}}$ approaches $\mathbf{0}$: when the sample size increases:

$$\mathsf{Var}\left[\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right] \xrightarrow{p} \mathbf{0}.$$

so that $\hat{\boldsymbol{\beta}}$ collapses to $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}.$$

GLM, OLS estimator is asymptotically not efficient

$\hat{\boldsymbol{\beta}}$ is asymptotically not efficient when

$$\operatorname{Var}\left[\boldsymbol{\varepsilon} \mid \boldsymbol{X}\right] = \sigma^2 \boldsymbol{\Omega}.$$

We do not prove this.

GLM, OLS estimator is asymptotically normal

Assuming that $\mathrm{E}[\boldsymbol{x}_i \varepsilon_i] = 0$ (A3), assuming that $\boldsymbol{x}_i$ is i.i.d. (A5), assuming that ε_i are uncorrelated, but allowing them to be heteroskedastic, and applying the CLT (Greene, Theorem D.19A) and hence not assuming that ε_i is normal (A6), it can be shown that

$$\hat{\boldsymbol{\beta}} \overset{a}{\sim} N\left[\beta, \frac{1}{n}\left(\mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{x}_i'\right]\right)^{-1} \mathrm{E}\left[\boldsymbol{x}_i \sigma^2 \omega_i \boldsymbol{x}_i'\right] \left(\mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{x}_i'\right]\right)^{-1}\right].$$

GLM, OLS estimator is asymptotically normal

Note that this result is for heteroskedasticity. The other type of nonspherical errors is where ε_i are correlated across i. Provided that this correlation diminishes with observations further away from each other, $\hat{\beta}$ is also asymptotically normal when errors are autocorrelated.

GLM, GLS estimator

$\hat{\boldsymbol{\beta}}_{\boldsymbol{OLS}}$ is not an efficient estimator in the GLM. The Generalized Least Squares (GLS) estimator,

$$\hat{\beta}_{GLS} = \left(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y},$$

is an alternative estimator which is efficient in the GLM. We do not study this estimator here. We note, however, that using this estimator is another way to deal with heteroskedasticity or autocorrelation in regression errors. In practice we do not use this estimator because it is computationally costly. To deal with non-spherical errors, we adapt another strategy, which we study next.

GLM, heteroskedasticity-consistent variance estimator

Consider the asymptotic variance of the OLS estimator shown above:

$$\text{Asy. Var}\left[\hat{\boldsymbol{\beta}}\right] = \frac{1}{n}\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}\mathrm{E}\left[\boldsymbol{x}_i\sigma^2\omega_i\boldsymbol{x}_i'\right]\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}.$$

The two expected value terms are unobserved since we do not have the information of the entire population. We do not observe σ^2. We do not know the form of $\boldsymbol{\Omega}$ and hence ω_i. We need a consistent estimator of this variance so that we can use it in practice. For what purpose we need this estimator is still to come.

GLM, heteroskedasticity-consistent variance estimator

With certain assumptions on $\boldsymbol{x}_i$, and using the LLN (Greene, Theorems D.4 through D.9), it can be shown that we can consistently estimate

$$\text{Asy. Var}\left[\hat{\boldsymbol{\beta}}\right] = \frac{1}{n}\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}\mathrm{E}\left[\boldsymbol{x}_i\sigma^2\omega_i\boldsymbol{x}_i'\right]\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}$$

with

$$\text{Est. Asy. Var}\left[\hat{\boldsymbol{\beta}}\right] = \frac{1}{n}\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2\boldsymbol{x}_i\boldsymbol{x}_i'\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}.$$

Dropping the $\frac{1}{n}$ terms,

$$\text{Est. Asy. Var}\left[\hat{\boldsymbol{\beta}}\right] = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_i^2\boldsymbol{x}_i\boldsymbol{x}_i'\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}.$$

GLM, heteroskedasticity-consistent variance estimator

$$\text{Est. Asy. Var}\left[\hat{\boldsymbol{\beta}}\right] = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 \boldsymbol{x}_i \boldsymbol{x}_i' \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

is called the heteroskedasticity-consistent variance estimator (HCVE) of the Asy. Var $\left[\hat{\boldsymbol{\beta}}\right]$. We said that the t and F statistics are not valid if we use

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

But they are valid if we use the HCVE. They are then called the heteroskedasticity-consistent t and F statistics. HCVE is powerful. $\boldsymbol{\Omega}$ is often unknown. HCVE does not need to figure out $\boldsymbol{\Omega}$. We can use the HCVE to make inference on $\boldsymbol{\beta}$. We only need to keep in mind that the HCVE, and the test statistics that make use of the HCVE, require that n is large. We also do not need to assume that the errors are normal!

GLM, heteroskedasticity-consistent variance estimator

We can recast

$$\text{Est. Asy. Var}\left[\hat{\boldsymbol{\beta}}\right] = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 \boldsymbol{x}_i \boldsymbol{x}_i' \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

as

$$\text{Est. Asy. Var}\left[\hat{\boldsymbol{\beta}}\right] = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}' diag\left(\hat{\varepsilon}_1^2, \ldots, \hat{\varepsilon}_n^2\right) \boldsymbol{X} \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1},$$

where

$$diag\left(\hat{\varepsilon}_1^2, \ldots, \hat{\varepsilon}_n^2\right) = \begin{bmatrix} \hat{\varepsilon}_1^2 & 0 & \ldots & 0 \\ 0 & \hat{\varepsilon}_2^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \hat{\varepsilon}_n^2 \end{bmatrix}.$$

GLM, heteroskedasticity-consistent variance estimator

How to interpret the HEC?

GLM, heteroskedasticity-consistent variance estimator

Start with the estimator of the Asy. Var $\left[\hat{\beta}\right]$ under homoskedasticity:

$$\begin{aligned}
\text{Est. Asy. Var}\left[\hat{\beta}\right] &= \hat{\sigma}^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n-K}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n-K}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n-K}\boldsymbol{I}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' diag\left(\frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n-K}, \ldots, \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n-K}\right)\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}.
\end{aligned}$$

We can move $\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$ across the matrices because it is a scalar.

GLM, heteroskedasticity-consistent variance estimator

Under homoskedasticity:

$$\text{Est. Asy. Var}\left[\hat{\boldsymbol{\beta}}\right] = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' diag\left(\frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n-K}, \ldots, \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n-K}\right)\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}.$$

Across the diagonal, the elements are the same!

Under heteroskedasticity:

$$\text{Est. Asy. Var}\left[\hat{\boldsymbol{\beta}}\right] = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' diag\left(\hat{\varepsilon}_1^2, \ldots, \hat{\varepsilon}_n^2\right)\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1},$$

Across the diagonal, the elements are different! You are accounting for heteroskedasticity!

GLM, heteroskedasticity-consistent variance estimator

```
. regress wage educ

      Source |       SS           df       MS      Number of obs   =       997
-------------+----------------------------------   F(1, 995)       =    251.46
       Model |  7842.35455         1  7842.35455   Prob > F        =    0.0000
    Residual |  31031.0745       995  31.1870095   R-squared       =    0.2017
-------------+----------------------------------   Adj R-squared   =    0.2009
       Total |   38873.429       996  39.0295472   Root MSE        =    5.5845

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   1.135645   .0716154    15.86   0.000     .9951106     1.27618
       _cons |  -4.860424   .9679821    -5.02   0.000    -6.759944   -2.960903
------------------------------------------------------------------------------
```

GLM, heteroskedasticity-consistent estimator

```
. regress wage educ, robust

Linear regression                               Number of obs     =        997
                                                F(1, 995)         =     178.66
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2017
                                                Root MSE          =     5.5845
```

wage	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	1.135645	.0849627	13.37	0.000	.9689186	1.302372
_cons	-4.860424	1.078429	-4.51	0.000	-6.976681	-2.744167

GLM, heteroskedasticity-consistent estimator

We can, and should always use the heteroskedasticity-consistent S.E. estimator because if there is no heteroskedasticity, it is equivalent to the standard S.E. estimator. We do not prove it.

GLS estimator or the HCVE?

If we detect heteroskedasticity, should we use the GLS estimator, which is a coefficient estimator, or the HCVE, which is a S.E. estimator?

In the GLS approach, we alter the model and hence the coefficient estimator that is a function of the altered model. We also alter the standard errors since they depend on the altered coefficient estimator.

The HCVE does not do anything to the coefficient estimator. It acknowledges heteroskedasticity and accounts for it in the standard error of the coefficient estimator.

The advantage of the HCVE is that we do not need to figure out the covariance structure of the errors. It is difficult to obtain a reasonable estimate of the error covariance structure, and hence to use the GLS estimator.