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

Econometrics (35B206), Lecture 4

Tunga Kantarcı, TiSEM, Tilburg University, Spring 2019

SLM, spherical errors assumption

ε_i is spherical if 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

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

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

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

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

$\boldsymbol{\varepsilon}$ is $n \times 1$. $\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'$ is $n \times n$. Hence, $\mathrm{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}]$ 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 a model is its assumptions. 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.

GLM

As we relax the spherical errors assumption, we can differentiate between two cases. First, if we relax the assumption that ε_i is homoskedastic, then ε_i is heteroskedastic. Second, if we relax the assumption that ε_i is non-autocorrelated, then ε_i is autocorrelated.

GLM, error is heteroskedastic

We start with heteroskedasticity.

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 $\omega(x_i)$. $\sigma^2 \omega(x_i)$ 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, $\boldsymbol{\varepsilon}$, the variance-covariance matrix is

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

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 $\operatorname{E}[\varepsilon\varepsilon' \mid \boldsymbol{X}]$ look like?

GLM, error is heteroskedastic

$$
\begin{aligned}
\mathsf{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}\right] &= \begin{bmatrix} \mathsf{E}\left[\varepsilon_1\varepsilon_1 \mid \boldsymbol{X}\right] & \mathsf{E}\left[\varepsilon_1\varepsilon_2 \mid \boldsymbol{X}\right] & \dots & \mathsf{E}\left[\varepsilon_1\varepsilon_n \mid \boldsymbol{X}\right] \\ \mathsf{E}\left[\varepsilon_2\varepsilon_1 \mid \boldsymbol{X}\right] & \mathsf{E}\left[\varepsilon_2\varepsilon_2 \mid \boldsymbol{X}\right] & \dots & \mathsf{E}\left[\varepsilon_2\varepsilon_n \mid \boldsymbol{X}\right] \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{E}\left[\varepsilon_n\varepsilon_1 \mid \boldsymbol{X}\right] & \mathsf{E}\left[\varepsilon_n\varepsilon_2 \mid \boldsymbol{X}\right] & \dots & \mathsf{E}\left[\varepsilon_n\varepsilon_n \mid \boldsymbol{X}\right] \end{bmatrix} \\
&= \underbrace{\begin{bmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \omega_n \end{bmatrix}}_{\boldsymbol{\Omega}} \sigma^2.
\end{aligned}
$$

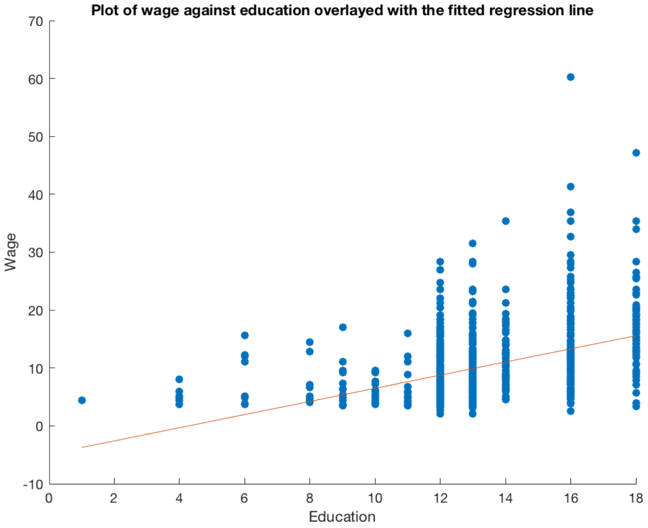
GLM, error is heteroskedastic

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

GLM, error is heteroskedastic, example

We want to explain wage with education.

GLM, error is heteroskedastic, example



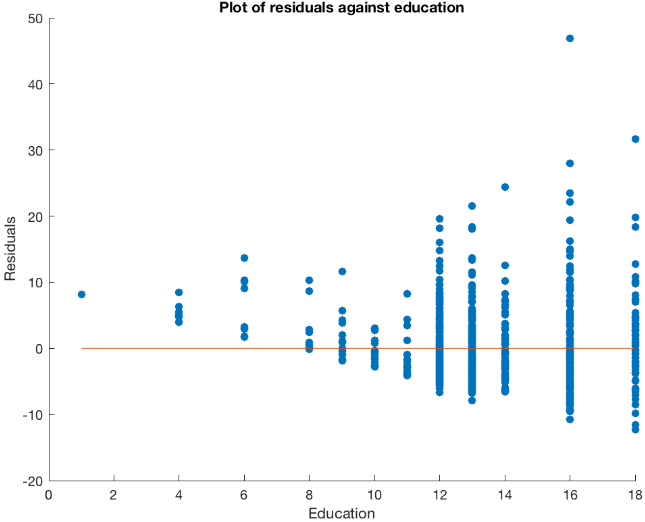
GLM, error is heteroskedastic, example

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 be more variable at higher levels of education. Hence, errors do not have a constant variance, at 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 the variance of the error, at given values of education, is not constant, then the variance of wage, 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,

$$\begin{aligned}
\text{Var}[y_i \mid x_i] &= \text{Var}[x_i\beta \mid x_i] + \text{Var}[\varepsilon_i \mid x_i] \\
&= \beta^2\text{Var}[x_i \mid x_i] + \text{Var}[\varepsilon_i \mid x_i] \\
&= \text{Var}[\varepsilon_i \mid x_i] \\
&= \sigma_i^2.
\end{aligned}$$

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.

GLM, error is autocorrelated

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

GLM, error is autocorrelated

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

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

GLM, error is autocorrelated

$\boldsymbol{\Omega}$ 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 $\upsilon_t \sim \mathit{IID}\left(0, \sigma_\upsilon^2\right)$, 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

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

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

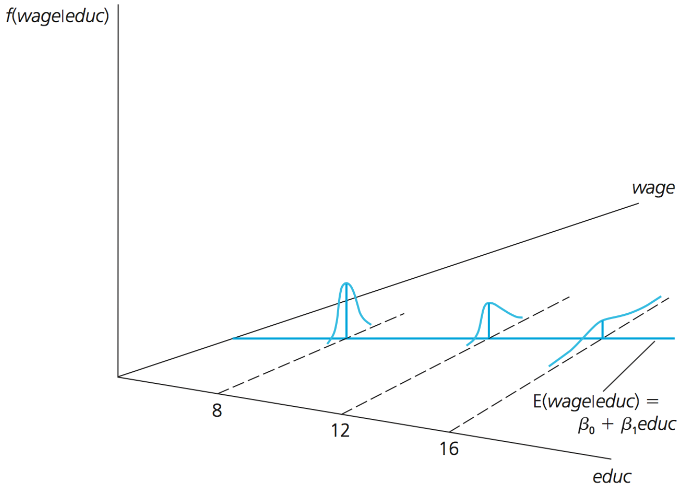
A3: Strict exogeneity: $\mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_k\right] = 0$. Hence, the conditional expectation function follows.

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? 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

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

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

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

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

GLM, OLS estimator is not efficient

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

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

since $\mathrm{E}\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta}$ 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,

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

and

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

We have shown that

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

where $\hat{\boldsymbol{\beta}}_0$ is a competitor 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,

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

and

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

Is

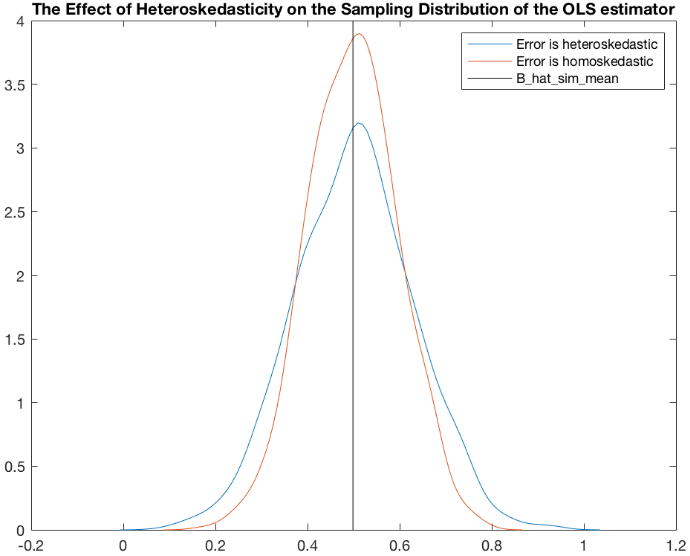
$$\text{Var}\left[\hat{\boldsymbol{\beta}}_0 \mid \boldsymbol{X}\right] \geq \text{Var}\left[\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right]$$

still true in the GLM? It is not! Later we will prove that

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

where $\hat{\boldsymbol{\beta}}_{GLS}$ is a new estimator. OLS estimator is not efficient in the GLM!

GLM, OLS estimator is not efficient



GLM, OLS estimator is not efficient, implications

$\hat{\beta}$ has a different 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

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

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 in these statistics! Can 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, if we use a consistent estimator of this variance. We will derive this estimator later.

GLM, OLS estimator is normal

We know that

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

Assume that ε 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 is 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,

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

GLM, OLS estimator is consistent

In the GLM,

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

and

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

GLM, OLS estimator is consistent

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

GLM, OLS estimator is consistent

$$\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} = \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1},$$

and

$$\frac{1}{n}\boldsymbol{X}'\Omega\boldsymbol{X} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\omega_i\boldsymbol{x}_i'.$$

GLM, OLS estimator is consistent

We know that

$$\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1} \xrightarrow{p} \left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}.$$

Similarly,

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\omega_i\boldsymbol{x}_i' \xrightarrow{p} \mathrm{E}\left[\boldsymbol{x}_i\omega_i\boldsymbol{x}_i'\right]$$

under certain assumptions on ω_i that we do not elaborate on.
Furthermore,

$$\frac{\sigma^2}{n} \to 0$$

as $n \to \infty$.

GLM, OLS estimator is consistent

Then,

$$\mathrm{Var}\left[\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right] = \underbrace{\frac{1}{n}}_{\to 0}\sigma^2 \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}}_{\xrightarrow{p}\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}} \underbrace{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\omega_i\boldsymbol{x}_i'}_{\xrightarrow{p}\mathrm{E}\left[\boldsymbol{x}_i\omega_i\boldsymbol{x}_i'\right]} \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}}_{\xrightarrow{p}\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}},$$

and therefore we have

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

This shows that in the GLM,

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

GLM, OLS estimator is asymptotically not efficient

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

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

We do not prove this.

GLM, OLS estimator is asymptotically normal

Consider our earlier result that

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) = \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1} \sqrt{n}\,\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \varepsilon_i .$$

GLM, OLS estimator is asymptotically normal

We already know that

$$\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1} \xrightarrow{p} \left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}.$$

Convergence in probability implies convergence in distribution. Hence,

$$\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1} \xrightarrow{d} \left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}.$$

$\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]$ is a constant matrix and has no variance.

GLM, OLS estimator is asymptotically normal

Assuming that $\mathrm{E}\left[\boldsymbol{x}_i \varepsilon_i\right] = 0$ (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),

$$\sqrt{n}\,\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \varepsilon_i \xrightarrow{d} N\left[\mathbf{0}, \mathrm{E}\left[\boldsymbol{x}_i \sigma^2 \omega_i \boldsymbol{x}_i'\right]\right].$$

We did not assume that ε_i is normal (A6). We are enjoying the CLT!

GLM, OLS estimator is asymptotically normal

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)=\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}}_{\xrightarrow{d}\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}}\underbrace{\sqrt{n}\,\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\varepsilon_i}_{\xrightarrow{d}N\left[\mathbf{0},\mathrm{E}\left[\boldsymbol{x}_i\sigma^2\omega_i\boldsymbol{x}_i'\right]\right]}\quad.$$

Using the product rule of limiting distributions (Greene, Theorem D.16),

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\xrightarrow{d}\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}N\left[\mathbf{0},\mathrm{E}\left[\boldsymbol{x}_i\sigma^2\omega_i\boldsymbol{x}_i'\right]\right].$$

GLM, OLS estimator is asymptotically normal

$$
\begin{aligned}
\sqrt{n}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) &\xrightarrow{d} \left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1} N\left[\mathbf{0}, \mathrm{E}\left[\boldsymbol{x}_i\sigma^2\omega_i\boldsymbol{x}_i'\right]\right] \\
&\xrightarrow{d} N\left[\mathbf{0}, \left(\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}\right)\mathrm{E}\left[\boldsymbol{x}_i\sigma^2\omega_i\boldsymbol{x}_i'\right]\left(\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}\right)'\right] \\
&\xrightarrow{d} N\left[\mathbf{0}, \left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}\mathrm{E}\left[\boldsymbol{x}_i\sigma^2\omega_i\boldsymbol{x}_i'\right]\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}\right],
\end{aligned}
$$

using the property that the transpose and inverse operations commute from the second to the third line.

GLM, OLS estimator is asymptotically normal

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} N\left[\mathbf{0}, \left(\mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{x}_i'\right]\right)^{-1} \mathrm{E}\left[\boldsymbol{x}_i \sigma^2 \omega_i \boldsymbol{x}_i'\right] \left(\mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{x}_i'\right]\right)^{-1}\right]$$

implies 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, heteroskedasticity-consistent estimator

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

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 estimator

We know that the first expected value

$$\left(\mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{x}_i'\right]\right)^{-1}$$

is equal to

$$\left(\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1},$$

which we can estimate with

$$\left(\frac{1}{n} \boldsymbol{X}' \boldsymbol{X}\right)^{-1}.$$

GLM, heteroskedasticity-consistent estimator

We know that the second expected value

$$\mathrm{E}\left[\boldsymbol{x}_i \sigma^2 \omega_i \boldsymbol{x}_i'\right]$$

is equal to

$$\text{plim} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \sigma^2 \omega_i \boldsymbol{x}_i'.$$

How can we estimate it?

GLM, heteroskedasticity-consistent estimator

With certain assumptions on $\boldsymbol{x}_i$, and using the LLN (Greene, Theorems D.4 through D.9),

$$\text{plim}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\sigma^2\omega_i\boldsymbol{x}_i' = \text{plim}\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2\boldsymbol{x}_i\boldsymbol{x}_i'.$$

Furthermore, since $\hat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$, $\hat{\varepsilon}_i$ $(= y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}})$ is a consistent estimator of ε_i. Hence,

$$\text{plim}\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2\boldsymbol{x}_i\boldsymbol{x}_i' = \text{plim}\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2\boldsymbol{x}_i\boldsymbol{x}_i'.$$

The last term can be estimated with

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2\boldsymbol{x}_i\boldsymbol{x}_i'.$$

GLM, heteroskedasticity-consistent estimator

These results mean that we can consistently estimate

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

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

GLM, heteroskedasticity-consistent estimator

To calculate the HCE in MATLAB, 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 estimator

How to interpret the HEC?

GLM, heteroskedasticity-consistent 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 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 estimator

```
. regress wage educ
```

Source	SS	df	MS
Model	7842.35455	1	7842.35455
Residual	31031.0745	995	31.1870095
Total	38873.429	996	39.0295472

Number of obs	=	997
F(1, 995)	=	251.46
Prob > F	=	0.0000
R-squared	=	0.2017
Adj R-squared	=	0.2009
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, GLS estimator

$\hat{\boldsymbol{\beta}}$ is an inefficient estimator in the GLM. We now derive a new estimator which is efficient in the GLM. In this derivation we assume that ε_i is heteroskedastic, but not autocorrelated.

GLM, GLS estimator

Our new estimator will make use of the particular matrix $\boldsymbol{\Omega}^{-1}$. As we have motivated already, $\boldsymbol{\Omega}$ represents the structural component of the variance of the error term in the GLM. That is,

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

Our derivation of the new estimator starts with a certain treatment of the $\boldsymbol{\Omega}^{-1}$ matrix.

GLM, GLS estimator

Assume that $\boldsymbol{\Omega}$ is a symmetric, positive definite matrix. Then, we can use the Cholesky decomposition (Greene, A.6.11) to obtain

$$\boldsymbol{\Omega}^{-1} = \boldsymbol{P}'\boldsymbol{P}.$$

It is also true that

$$\boldsymbol{\Omega} = \boldsymbol{P}^{-1}\left(\boldsymbol{P}^{-1}\right)'.$$

$\boldsymbol{\Omega}^{-1}$ is $n \times n$. $\boldsymbol{P}$ is $n \times n$. How do they look like?

GLM, GLS estimator

$$\mathbf{\Omega} = \begin{bmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & \omega_n \end{bmatrix},$$

and

$$\mathbf{\Omega}^{-1} = \begin{bmatrix} \frac{1}{\omega_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\omega_2} & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & \frac{1}{\omega_n} \end{bmatrix}.$$

GLM, GLS estimator

Using the Cholesky decomposition,

$$\boldsymbol{\Omega}^{-1} = \boldsymbol{P}'\boldsymbol{P}.$$

Hence,

$$\boldsymbol{\Omega}^{-1} = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{\omega_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{\omega_2}} & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & \frac{1}{\sqrt{\omega_n}} \end{bmatrix}}_{\boldsymbol{P}'} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{\omega_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{\omega_2}} & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & \frac{1}{\sqrt{\omega_n}} \end{bmatrix}}_{\boldsymbol{P}}.$$

GLM, GLS estimator

Premultiply the liner model with $\boldsymbol{P}$ to obtain

$$\boldsymbol{P}\boldsymbol{y} = \boldsymbol{P}\boldsymbol{X}\beta + \boldsymbol{P}\varepsilon.$$

What is the variance of the transformed error, conditional on $\boldsymbol{X}$?

GLM, GLS estimator

If $\mathrm{E}[\varepsilon \mid \boldsymbol{X}] = 0$, and because the transpose and inverse operators commute,

$$\begin{aligned}
\mathrm{Var}[\boldsymbol{P}\varepsilon \mid \boldsymbol{X}] &= \mathrm{E}\left[\boldsymbol{P}\varepsilon\varepsilon'\boldsymbol{P}' \mid \boldsymbol{X}\right] - \mathrm{E}[\boldsymbol{P}\varepsilon \mid \boldsymbol{X}]\left(\mathrm{E}[\boldsymbol{P}\varepsilon \mid \boldsymbol{X}]\right)' \\
&= \boldsymbol{P}\mathrm{E}\left[\varepsilon\varepsilon' \mid \boldsymbol{X}\right]\boldsymbol{P}' - \boldsymbol{P}\mathrm{E}[\varepsilon \mid \boldsymbol{X}]\left(\boldsymbol{P}\mathrm{E}[\varepsilon \mid \boldsymbol{X}]\right)' \\
&= \boldsymbol{P}\sigma^2\boldsymbol{\Omega}\boldsymbol{P}' \\
&= \sigma^2\boldsymbol{P}\boldsymbol{P}^{-1}\left(\boldsymbol{P}^{-1}\right)'\boldsymbol{P}' \\
&= \sigma^2\boldsymbol{P}\boldsymbol{P}^{-1}\left(\boldsymbol{P}'\right)^{-1}\boldsymbol{P}' \\
&= \sigma^2\boldsymbol{I}.
\end{aligned}$$

The transformed error is now homoskedastic! This is the idea of GLS. We do nothing but weigh each observation of the data with the inverse of the square root of the problematic structure so that the conditional variance of the error is cleared from that structure.

GLM, GLS estimator

In the transformed model

$$\boldsymbol{P}\boldsymbol{y} = \boldsymbol{P}\boldsymbol{X}\beta + \boldsymbol{P}\boldsymbol{\varepsilon},$$

$\mathrm{Var}\left[\boldsymbol{P}\boldsymbol{\varepsilon} \mid \boldsymbol{X}\right] = \sigma^2 \boldsymbol{I}$. We can estimate the model using the OLS estimator! Apply the OLS estimator on the transformed model!

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{GLS} &= \left((\boldsymbol{P}\boldsymbol{X})' \boldsymbol{P}\boldsymbol{X}\right)^{-1} (\boldsymbol{P}\boldsymbol{X})' \boldsymbol{P}\boldsymbol{y} \\
&= \left(\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{y} \\
&= \left(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}.
\end{aligned}$$

$\hat{\boldsymbol{\beta}}_{GLS}$ is the generalised least squares estimator of $\boldsymbol{\beta}$. It is an efficient estimator because the homoskedasticity assumption holds in the transformed model! $\hat{\boldsymbol{\beta}}_{GLS}$ differs from $\hat{\boldsymbol{\beta}}$ in the weighting matrix, which is $\boldsymbol{\Omega}^{-1}$ for $\hat{\boldsymbol{\beta}}_{GLS}$, while it is $\boldsymbol{I}$ for $\hat{\boldsymbol{\beta}}$.

GLM, GLS estimator is unbiased

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{GLS} &= \left(\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{y} \\
&= \left(\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{P}'\left(\boldsymbol{P}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{P}\boldsymbol{\varepsilon}\right) \\
&= \boldsymbol{\beta} + \left(\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{\varepsilon}.
\end{aligned}$$

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

$$\mathrm{E}\left[\hat{\boldsymbol{\beta}}_{GLS} \mid \boldsymbol{X}\right] = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P}\,\mathrm{E}\left[\boldsymbol{\varepsilon} \mid \boldsymbol{X}\right].$$

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

$$\mathrm{E}\left[\hat{\boldsymbol{\beta}}_{GLS} \mid \boldsymbol{X}\right] = \boldsymbol{\beta}.$$

GLM, GLS estimator is efficient

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{GLS} &= \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{\varepsilon} \\ &= \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}.\end{aligned}$$

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

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

GLM, GLS estimator is efficient

It can be shown that

$$\mathrm{Var}\left[\hat{\boldsymbol{\beta}}_0 \mid \boldsymbol{X}\right] - \mathrm{Var}\left[\hat{\boldsymbol{\beta}}_{GLS} \mid \boldsymbol{X}\right] \geq 0.$$

We do not prove this. It says that $\mathrm{Var}\left[\hat{\boldsymbol{\beta}}_{GLS} \mid \boldsymbol{X}\right]$ is the smallest when compared to the variance of any other linear unbiased estimator. Hence, $\hat{\boldsymbol{\beta}}_{GLS}$ is efficient.

GLM, GLS estimator is normal

Suppose that $\boldsymbol{\varepsilon}$ is multivariate normal. If we fix $\boldsymbol{X}$, $\hat{\boldsymbol{\beta}}_{GLS}$ is also multivariate normal, since $\hat{\boldsymbol{\beta}}_{GLS}$ is a linear function of $\boldsymbol{\varepsilon}$. Then, using the mean vector and the variance-covariance matrix of $\hat{\boldsymbol{\beta}}_{GLS}$ derived above,

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

GLM, GLS estimator is consistent

$$\begin{aligned}
\mathrm{Var}\left[\hat{\boldsymbol{\beta}}_{GLS} \mid \boldsymbol{X}\right] &= \sigma^2 \left(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right)^{-1} \\
&= \frac{1}{n}\sigma^2 \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right)^{-1}.
\end{aligned}$$

$$\frac{\sigma^2}{n} \to 0$$

as $n \to \infty$. And

$$\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right)^{-1} = \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\omega_i^{-1}\boldsymbol{x}_i'\right)^{-1} \xrightarrow{p} \mathrm{E}\left[\boldsymbol{x}_i\omega_i^{-1}\boldsymbol{x}_i'\right].$$

Hence,

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

GLM, GLS estimator is asymptotically normal

It can be shown that

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{GLS} - \boldsymbol{\beta}\right) \xrightarrow{d} N\left[\mathbf{0}, \sigma^2 \left(\mathrm{E}\left[\boldsymbol{x}_i \omega_i^{-1} \boldsymbol{x}_i'\right]\right)^{-1}\right],$$

which implies that

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

σ^2 and $1/n \left(\mathrm{E}\left[\omega_i^{-1} \boldsymbol{x}_i \boldsymbol{x}_i'\right]\right)^{-1}$ are unobserved. They are estimated with $\hat{\sigma}^2$ and $\left(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right)^{-1}$ given the sample data at hand, respectively.

GLM, FGLS estimator

We can employ

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

if $\boldsymbol{\Omega}$ is known. If $\boldsymbol{\Omega}$ depends on unknown parameters, we cannot compute $\hat{\boldsymbol{\beta}}_{GLS}$. However, if $\boldsymbol{\Omega}$ depends on a small set of unknown parameters, we can estimate them, and construct $\hat{\boldsymbol{\Omega}}$. We can then employ the estimator

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

This is the feasible generalised least squares estimator.

GLM, FGLS estimator is consistent

It can be shown that

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

GLM, FGLS estimator is asymptotically efficient

It can be shown that $\hat{\beta}_{FGLS}$ is asymptotically efficient.

GLM, FGLS estimator is asymptotically normal

It can be shown that

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

σ^2 and $1/n\left(\mathrm{E}\left[\omega_i^{-1}\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}$ are unobserved. They are estimated with $\hat{\sigma}^2$ and $\left(\boldsymbol{X}'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{X}\right)^{-1}$ given the sample data at hand, respectively.

GLM, FGLS estimator

Consider

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

and

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

Under certain regularity conditions, and using asymptotic theory, it can be shown that in many situations

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{FGLS} - \hat{\boldsymbol{\beta}}_{GLS}\right) \xrightarrow{p} 0.$$

This implies that asymptotically, we do not lose anything when we replace $\boldsymbol{\Omega}$ by $\hat{\boldsymbol{\Omega}}$. In a large sample, $\hat{\boldsymbol{\beta}}_{FGLS}$ behaves approximately in the same way as $\hat{\boldsymbol{\beta}}_{GLS}$. $\hat{\boldsymbol{\beta}}_{GLS}$ cannot be computed, but it has good statistical properties. $\hat{\boldsymbol{\beta}}_{FGLS}$ can be computed, and it has approximately the same properties.

GLM, FGLS estimator, example

Consider the linear model

$$y_i = \boldsymbol{x}_i'\boldsymbol{\gamma} + \varepsilon_i.$$

A popular way of specifying the form of heteroskedasticity is exponential, or multiplicative, heteroskedasticity. That is,

$$\varepsilon_i = e^{0.5\boldsymbol{x}_i'\boldsymbol{\beta}}\upsilon_i,\ \upsilon_i \mid \boldsymbol{x}_i' \sim N(0,1),$$

assuming that $\boldsymbol{x}_i'$ includes an intercept. Then,

$$\operatorname{Var}\left[\varepsilon_i \mid \boldsymbol{x}_i'\right] = \operatorname{E}\left[\varepsilon_i^2 \mid \boldsymbol{x}_i'\right] = e^{\boldsymbol{x}_i'\boldsymbol{\beta}}.$$

Checkpoint. This shows that $\boldsymbol{\Omega}$ depends on the K dimensional parameter vector $\boldsymbol{\beta}$. We can estimate $\boldsymbol{\beta}$, and apply FGLS. How do we estimate $\boldsymbol{\beta}$?

GLM, FGLS estimator, example

$$\operatorname{Var}\left[\varepsilon_i \mid \boldsymbol{x}_i'\right] = \operatorname{E}\left[\varepsilon_i^2 \mid \boldsymbol{x}_i'\right] = e^{\boldsymbol{x}_i'\boldsymbol{\beta}}.$$

This conditional expectation function calls for a regression of ε_i^2 on $e^{\boldsymbol{x}_i'\boldsymbol{\beta}}$. This regression is nonlinear in the coefficients, and estimation is difficult. Instead, we can first linearise ε_i^2 by taking the logarithm of it, and then by taking the conditional expectation of $\ln\left(\varepsilon_i^2\right)$ to obtain

$$\operatorname{E}\left[\ln\left(\varepsilon_i^2\right) \mid \boldsymbol{x}_i'\right] = \boldsymbol{x}_i'\boldsymbol{\beta} + \operatorname{E}\left[\ln\left(\upsilon_i^2\right)\right].$$

GLM, FGLS estimator, example

$$\mathrm{E}\left[\ln\left(\varepsilon_i^2\right) \mid \boldsymbol{x}_i'\right] = \boldsymbol{x}_i'\boldsymbol{\beta} + \mathrm{E}\left[\ln\left(\upsilon_i^2\right)\right].$$

This conditional expectation function calls for a regression of $\ln\left(\varepsilon_i^2\right)$ on $\boldsymbol{x}_i'$. ε_i is not observed but it can be estimated with the OLS residual $\hat{\varepsilon}_i$ from the regression

$$y_i = \boldsymbol{x}_i'\boldsymbol{\gamma} + \varepsilon_i.$$

$\mathrm{E}\left[\ln\left(\upsilon_i^2\right)\right]$ is some constant and it does not depend on $\boldsymbol{x}_i'$.

GLM, FGLS estimator, example

$$\mathrm{E}\left[\ln\left(\varepsilon_i^2\right) \mid \boldsymbol{x}_i'\right] = \boldsymbol{x}_i'\boldsymbol{\beta} + \mathrm{E}\left[\ln\left(\upsilon_i^2\right)\right].$$

The OLS estimator gives consistent estimates of all slope coefficients in $\boldsymbol{\beta}$. However, $\boldsymbol{\beta}_0$ in $\boldsymbol{\beta}$ and $\mathrm{E}\left[\ln\left(\upsilon_i^2\right)\right]$ make up the constant term of the regression equation, and therefore, as it can be shown, the OLS estimator of $\boldsymbol{\beta}_0$ is not consistent. However, this does not matter since it is good enough to approximate $\boldsymbol{\Omega}$ up to some scalar not depending on $\boldsymbol{x}_i'$.

GLM, FGLS estimator, example

$$\mathrm{E}\left[\ln\left(\varepsilon_i^2\right) \mid \boldsymbol{x}_i'\right] = \boldsymbol{x}_i'\boldsymbol{\beta} + \mathrm{E}\left[\ln\left(\upsilon_i^2\right)\right].$$

In the regression of $\ln\left(\hat{\varepsilon}_i^2\right)$ on $\boldsymbol{x}_i'$, joint significance of the coefficients provides statistical evidence that the errors of the assumed model are heteroskedastic in terms of the given form.

GLM, FGLS estimator, example

Summary of the steps of the FGLS regression for this example:

1. Estimate the LRM to obtain the residual vector $\hat{\boldsymbol{\varepsilon}}$.

2. Square each term of the vector $\hat{\boldsymbol{\varepsilon}}$. Use the squared residuals as estimators of the elements on the diagonal of $\text{Var}\left[\varepsilon \mid \boldsymbol{X}\right]$. Take the logarithm of this vector.

3. Treat this vector as the dependent variable, and $\boldsymbol{X}$ as the matrix of independent variables. Carry out OLS regression. $\hat{\boldsymbol{\beta}}$ is a consistent estimate of $\boldsymbol{\beta}$. Obtain the prediction vector $\boldsymbol{X}\hat{\boldsymbol{\beta}}$.

4. Calculate $e^{\boldsymbol{X}\hat{\boldsymbol{\beta}}}$. Construct $\hat{\boldsymbol{\Omega}}$ so that the elements of $e^{\boldsymbol{X}\hat{\boldsymbol{\beta}}}$ are on the diagonal of $\hat{\boldsymbol{\Omega}}$.

5. Calculate $\hat{\boldsymbol{\beta}}_{FGLS} = \left(\boldsymbol{X}'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{y}$.

GLM, FGLS estimator, example

Editor – /Users/Tunga/Library/Mobile Documents/com~ap

example.m | examplefunctionlss.m | +

```matlab
% FGLS regression example

clear;
load '/Users/Tunga/Library/Mobile Documents/c
y = unaid;
N = length(y);
X = [ones(N,1) dur ncb rank year];
lss = examplefunctionlss(y,X);
e_hat = lss.u_hat;
e_hat_sq_log = log(e_hat.^2);
X = [ones(N,1) dur ncb rank year];
lss_hk = examplefunctionlss(e_hat_sq_log,X);
O = diag(exp(lss_hk.y_hat));
OI = inv(O);
P = chol(OI,'lower');
y_t = P'*y;
X_t = P'*X;
lss_fgls = examplefunctionlss(y_t,X_t);
B_hat_fgls = lss_fgls.B_hat;
```

GLM, FGLS estimator, example

O

2677x2677 double

	1	2	3	4	5	6	7	8	9	10	1
1	31.0372	0	0	0	0	0	0	0	0	0	
2	0	28.2850	0	0	0	0	0	0	0	0	
3	0	0	34.5710	0	0	0	0	0	0	0	
4	0	0	0	42.2539	0	0	0	0	0	0	
5	0	0	0	0	33.2506	0	0	0	0	0	
6	0	0	0	0	0	28.1580	0	0	0	0	
7	0	0	0	0	0	0	27.6150	0	0	0	
8	0	0	0	0	0	0	0	31.3639	0	0	
9	0	0	0	0	0	0	0	0	30.7591	0	
10	0	0	0	0	0	0	0	0	0	41.8604	
11	0	0	0	0	0	0	0	0	0	0	41
12	0	0	0	0	0	0	0	0	0	0	
13	0	0	0	0	0	0	0	0	0	0	
14	0	0	0	0	0	0	0	0	0	0	
15	0	0	0	0	0	0	0	0	0	0	
16	0	0	0	0	0	0	0	0	0	0	
17	0	0	0	0	0	0	0	0	0	0	
18	0	0	0	0	0	0	0	0	0	0	
19	0	0	0	0	0	0	0	0	0	0	
20	0	0	0	0	0	0	0	0	0	0	
21	0	0	0	0	0	0	0	0	0	0	
22	0	0	0	0	0	0	0	0	0	0	

GLM, FGLS estimator, example

OI

2677x2677 double

	1	2	3	4	5	6	7	8	9	10	11
1	0.0322	0	0	0	0	0	0	0	0	0	
2	0	0.0354	0	0	0	0	0	0	0	0	
3	0	0	0.0289	0	0	0	0	0	0	0	
4	0	0	0	0.0237	0	0	0	0	0	0	
5	0	0	0	0	0.0301	0	0	0	0	0	
6	0	0	0	0	0	0.0355	0	0	0	0	
7	0	0	0	0	0	0	0.0362	0	0	0	
8	0	0	0	0	0	0	0	0.0319	0	0	
9	0	0	0	0	0	0	0	0	0.0325	0	
10	0	0	0	0	0	0	0	0	0	0.0239	
11	0	0	0	0	0	0	0	0	0	0	0.
12	0	0	0	0	0	0	0	0	0	0	
13	0	0	0	0	0	0	0	0	0	0	
14	0	0	0	0	0	0	0	0	0	0	
15	0	0	0	0	0	0	0	0	0	0	
16	0	0	0	0	0	0	0	0	0	0	
17	0	0	0	0	0	0	0	0	0	0	
18	0	0	0	0	0	0	0	0	0	0	
19	0	0	0	0	0	0	0	0	0	0	
20	0	0	0	0	0	0	0	0	0	0	
21	0	0	0	0	0	0	0	0	0	0	
22	0	0	0	0	0	0	0	0	0	0	

GLM, FGLS estimator, example

P

2677x2677 double

	1	2	3	4	5	6	7	8	9	10	11	1
1	0.1795	0	0	0	0	0	0	0	0	0	0	
2	0	0.1880	0	0	0	0	0	0	0	0	0	
3	0	0	0.1701	0	0	0	0	0	0	0	0	
4	0	0	0	0.1538	0	0	0	0	0	0	0	
5	0	0	0	0	0.1734	0	0	0	0	0	0	
6	0	0	0	0	0	0.1885	0	0	0	0	0	
7	0	0	0	0	0	0	0.1903	0	0	0	0	
8	0	0	0	0	0	0	0	0.1786	0	0	0	
9	0	0	0	0	0	0	0	0	0.1803	0	0	
10	0	0	0	0	0	0	0	0	0	0.1546	0	
11	0	0	0	0	0	0	0	0	0	0	0.1561	
12	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	
14	0	0	0	0	0	0	0	0	0	0	0	
15	0	0	0	0	0	0	0	0	0	0	0	
16	0	0	0	0	0	0	0	0	0	0	0	
17	0	0	0	0	0	0	0	0	0	0	0	
18	0	0	0	0	0	0	0	0	0	0	0	
19	0	0	0	0	0	0	0	0	0	0	0	
20	0	0	0	0	0	0	0	0	0	0	0	
21	0	0	0	0	0	0	0	0	0	0	0	

GLM, FGLS, econometric application: RE estimator

Consider the panel LRM

$$y_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta} + \mu_i + \nu_{it}.$$

y_{it}: an observation for individual i at time t.

$\boldsymbol{x}'_{it}$: K independent variables for individual i at time t. $1 \times K$. A variable can be time-varying or time-invariant.

$\boldsymbol{\beta}$: vector of true coefficients. $K \times 1$.

GLM, FGLS, econometric application: RE estimator

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \mu_i + \nu_{it}.$$

μ_i: time-invariant error, specific to i. $\mu_i \sim N(0, \sigma_\mu^2)$, and hence it is a random effect. Hence, the model is called the random effects model. μ_i captures individual heterogeneity. μ_i is assumed to be uncorrelated with $\boldsymbol{x}_{it}$ which is a strong assumption. μ_i are uncorrelated among individuals.

ν_{it}: time-variant error. $\nu_{it} \sim N(0, \sigma_\nu^2)$. ν_{it} is uncorrelated with X_{it}. ν_{it} are uncorrelated among individuals, but also within individuals.

ε_{it}: define it as $\mu_i + \nu_{it}$. It is the composite error of the model. μ_i and ν_{it} are independent.

GLM, FGLS, econometric application: RE estimator

Derive the variance-covariance matrix of ε_{it}. That is, derive $\mathrm{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}\right]$.

GLM, FGLS, econometric application: RE estimator

The variance for i, at a given t:

$$\mathrm{E}\left[\varepsilon_{it}\varepsilon_{it} \mid \boldsymbol{x}'_{it}\right] = \sigma_\mu^2 + \sigma_\nu^2.$$

The covariance for i, across t: ε_{it} are correlated within an individual, due to the time-invariant μ_i. Serial correlation! That is, for every i, and $t \neq s$,

$$\mathrm{E}\left[\varepsilon_{it}\varepsilon_{is} \mid \boldsymbol{x}'_{it}\right] = \sigma_\mu^2.$$

The covariance for i and j, across t: ε_{it} are not correlated across individuals. That is, for every $i \neq j$,

$$\mathrm{E}\left[\varepsilon_{it}\varepsilon_{js} \mid \boldsymbol{x}'_{it}\right] = 0.$$

GLM, FGLS, econometric application: RE estimator

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

GLM, FGLS, econometric application: RE estimator

For T errors of individual i, stored in $\boldsymbol{\varepsilon}_i$,

$$
\begin{aligned}
\operatorname{E}\left[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' \mid \boldsymbol{X}_i'\right] &= \begin{bmatrix}
\sigma_\mu^2+\sigma_\nu^2 & \sigma_\mu^2 & \cdots & \sigma_\mu^2 & \cdots & \sigma_\mu^2 \\
\sigma_\mu^2 & \sigma_\mu^2+\sigma_\nu^2 & \cdots & \sigma_\mu^2 & \cdots & \sigma_\mu^2 \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\sigma_\mu^2 & \sigma_\mu^2 & \cdots & \sigma_\mu^2+\sigma_\nu^2 & \cdots & \sigma_\mu^2 \\
\vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\
\sigma_\mu^2 & \sigma_\mu^2 & \cdots & \sigma_\mu^2 & \cdots & \sigma_\mu^2+\sigma_\nu^2
\end{bmatrix}_{T\times T} \\
&\equiv \boldsymbol{\omega}_i
\end{aligned}
$$

ε_{it} are serially correlated within an individual!

GLM, FGLS, econometric application: RE estimator

For T errors of N individuals, stored in ε,

$$\mathrm{E}\left[\varepsilon\varepsilon' \mid \boldsymbol{X}\right] = \boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\omega}_1 & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\omega}_2 & \dots & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\omega}_n & \dots & \mathbf{0} \\ \vdots & \vdots & \dots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \boldsymbol{\omega}_N \end{bmatrix}_{NT \times NT}$$

GLM, FGLS, econometric application: RE estimator

Example $\mathbf{\Omega}$ for N = 2, T = 2,

$$\mathrm{E}\left[\varepsilon\varepsilon' \mid \boldsymbol{X}\right] = \mathbf{\Omega} = \begin{bmatrix} \sigma_\mu^2 + \sigma_\nu^2 & \sigma_\mu^2 & 0 & 0 \\ \sigma_\mu^2 & \sigma_\mu^2 + \sigma_\nu^2 & 0 & 0 \\ 0 & 0 & \sigma_\mu^2 + \sigma_\nu^2 & \sigma_\mu^2 \\ 0 & 0 & \sigma_\mu^2 & \sigma_\mu^2 + \sigma_\nu^2 \end{bmatrix}_{4\times 4}$$

GLM, FGLS, econometric application: RE estimator

ε_{it} are serially correlated within individuals. We cannot estimate $\boldsymbol{\beta}$ efficiently using OLS! What can we do?

GLM, FGLS, econometric application: RE estimator

We have constructed

$$\boldsymbol{\Omega}.$$

It includes the unknown σ_μ^2 and σ_ν^2. There are different ways to estimate them. We do not cover this. Use these estimates to obtain

$$\hat{\boldsymbol{\Omega}}.$$

Now GLS estimation is feasible. Then obtain

$$\hat{\boldsymbol{\Omega}}^{-1} = \hat{\boldsymbol{P}}'\hat{\boldsymbol{P}}.$$

GLM, FGLS, econometric application: RE estimator

$\hat{\boldsymbol{P}}\varepsilon$ is serially uncorrelated within individuals!

GLM, FGLS, econometric application: RE estimator

Obtain $\hat{\boldsymbol{P}}\boldsymbol{y}$ and $\hat{\boldsymbol{P}}\boldsymbol{X}$. Applying the OLS on them,

$$\begin{aligned}\hat{\beta}_{FGLS} &= (\underbrace{(\hat{\boldsymbol{P}}\boldsymbol{X})'}_{\boldsymbol{X}^{*\prime}}\underbrace{(\hat{\boldsymbol{P}}\boldsymbol{X})}_{\boldsymbol{X}^{*}})^{-1}\underbrace{(\hat{\boldsymbol{P}}\boldsymbol{X})'}_{\boldsymbol{X}^{*\prime}}\underbrace{\hat{\boldsymbol{P}}\boldsymbol{y}}_{\boldsymbol{y}^{*\prime}} \\ &= \hat{\boldsymbol{\beta}}_{RE}\end{aligned}$$

The feasible generalised least squares estimator is the random effects estimator!

GLM, RE estimator, example

Example in MATLAB, where we check how the $\hat{\boldsymbol{P}}$ matrix looks like.

GLM, RE estimator, example

Editor – /Users/Tunga/Library/Mobile Documents/com~apple~CloudDocs/Academic/Te

exercisepartwo.m | exerciseparttwofunctionbgs.m | exerciseparttwofunctionfes.m

```
% The RE estimator as the FGLS estimator

% 1. Load the data
clear;
filename = 'C:\Users\username\Desktop\exercise.csv';
delimiterIn = ',';
headerlinesIn = 1;
exercisedata = importdata(filename,delimiterIn,headerlinesIn);

% 2. Create the systematic component of the regression
y = exercisedata.data(:,1);
N = length(y);
X = [ones(N,1) exercisedata.data(:,2:end)];

% 3. Create additional variables
T = 8;
M = N/T;
P_D = kron(eye(M),ones(T,T).*1/T);

% 4. Obtain the RE coefficient estimates as FGLS coefficient estimates
fes = exerciseparttwofunctionfes(y,X,M,P_D,N);
bgs = exerciseparttwofunctionbgs(y,X,M,P_D,T);
s2_m = fes.s2_hat;
s2_v = bgs.s2_hat-s2_m/T;
L = 1-(T*s2_v/s2_m+1)^(-1/2);
P_hat = eye(N)-L*P_D;
y_t = P_hat*y;
X_t = P_hat*X;
B_hat = (X_t'*X_t)\(X_t'*y_t);
```

GLM, RE estimator, example

Editor – exercisepartwo.m | Variables – X_t

X_t

4360x10 double

	1	2	3	4	5	6	7	8	9
1	0.3634	0	-0.0796	0	0	5.0875	-1.8647	-15.2334	0
2	0.3634	0	0.9204	0	0	5.0875	-0.8647	-12.2334	0
3	0.3634	0	-0.0796	0	0	5.0875	0.1353	-7.2334	0
4	0.3634	0	-0.0796	0	0	5.0875	1.1353	-0.2334	0
5	0.3634	0	-0.0796	0	0	5.0875	2.1353	8.7666	0
6	0.3634	0	-0.0796	0	0	5.0875	3.1353	19.7666	0
7	0.3634	0	-0.0796	0	0	5.0875	4.1353	32.7666	0
8	0.3634	0	-0.0796	0	0	5.0875	5.1353	47.7666	0
9	0.3634	0	0	0	0	4.7241	-0.7745	-23.1512	0
10	0.3634	0	0	0	0	4.7241	0.2255	-14.1512	0
11	0.3634	0	0	0	0	4.7241	1.2255	-3.1512	0
12	0.3634	0	0	0	0	4.7241	2.2255	9.8488	0
13	0.3634	0	0	0	0	4.7241	3.2255	24.8488	0
14	0.3634	0	0	0	0	4.7241	4.2255	41.8488	0
15	0.3634	0	0	0	0	4.7241	5.2255	60.8488	0
16	0.3634	0	0	0	0	4.7241	6.2255	81.8488	0
17	0.3634	0	0	0.3634	0	4.3607	-0.7745	-23.1512	0
18	0.3634	0	0	0.3634	0	4.3607	0.2255	-14.1512	0
19	0.3634	0	0	0.3634	0	4.3607	1.2255	-3.1512	0
20	0.3634	0	0	0.3634	0	4.3607	2.2255	9.8488	0
21	0.3634	0	0	0.3634	0	4.3607	3.2255	24.8488	0
22	0.3634	0	0	0.3634	0	4.3607	4.2255	41.8488	0
23	0.3634	0	0	0.3634	0	4.3607	5.2255	60.8488	0
24	0.3634	0	0	0.3634	0	4.3607	6.2255	81.8488	0

GLM, RE estimator, example

Editor – exercisepartwo.m | Variables – P_hat

P_hat

4360x4360 double

	1	2	3	4	5	6	7	8	9
1	0.9204	-0.0796	-0.0796	-0.0796	-0.0796	-0.0796	-0.0796	-0.0796	0
2	-0.0796	0.9204	-0.0796	-0.0796	-0.0796	-0.0796	-0.0796	-0.0796	0
3	-0.0796	-0.0796	0.9204	-0.0796	-0.0796	-0.0796	-0.0796	-0.0796	0
4	-0.0796	-0.0796	-0.0796	0.9204	-0.0796	-0.0796	-0.0796	-0.0796	0
5	-0.0796	-0.0796	-0.0796	-0.0796	0.9204	-0.0796	-0.0796	-0.0796	0
6	-0.0796	-0.0796	-0.0796	-0.0796	-0.0796	0.9204	-0.0796	-0.0796	0
7	-0.0796	-0.0796	-0.0796	-0.0796	-0.0796	-0.0796	0.9204	-0.0796	0
8	-0.0796	-0.0796	-0.0796	-0.0796	-0.0796	-0.0796	-0.0796	0.9204	0
9	0	0	0	0	0	0	0	0	0.9204
10	0	0	0	0	0	0	0	0	-0.0796
11	0	0	0	0	0	0	0	0	-0.0796
12	0	0	0	0	0	0	0	0	-0.0796
13	0	0	0	0	0	0	0	0	-0.0796
14	0	0	0	0	0	0	0	0	-0.0796
15	0	0	0	0	0	0	0	0	-0.0796
16	0	0	0	0	0	0	0	0	-0.0796
17	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0

Testing for heteroskedasticity

In order to know whether to use the usual standard errors or the robust standard errors, or whether to use the FGLS estimator, it is useful to test for heteroskedasticity. Two tests are widely used. Both are not exact tests. That is, they are approximately valid in large samples. The tests are based on the same idea. If there is homoskedasticity, $\text{Var}\left[\varepsilon_i \mid \boldsymbol{x}_i\right]$ does not depend on $\boldsymbol{x}_i$, or on any function of $\boldsymbol{x}_i$. The OLS residuals $\hat{\varepsilon}_i$ can be used as approximations of the unobservable ε_i. An auxiliary regression can then be used to investigate whether $\hat{\varepsilon}_i^2$ is related to functions of $\boldsymbol{x}_i$.

Testing for het., White test of het. of unknown form

The null hypothesis we test is

$$H_0 : \sigma_i^2 = \sigma^2 \,\forall\, i.$$

The test consists of the following steps. 1. Compute $\hat{\boldsymbol{\beta}}$, and construct $\hat{\boldsymbol{\varepsilon}}$. 2. Square each term of the vector $\hat{\boldsymbol{\varepsilon}}$. Treat this vector as the dependent variable. Consider $\boldsymbol{X}$ as the matrix of all regressors and all their squares and interaction terms. Hence, $\boldsymbol{X}$ is a very general function of the regressors. Carry out OLS regression. 3. Compute $LM = nR^2$. The test statistic has an asymptotic χ^2 distribution with $J - 1$ degrees of freedom where J is the number of regressors including the column of ones.

Testing for het., Breusch-Pagan test of het. of known form

The null hypothesis of the test is

$$H_0 : \sigma_i^2 = \sigma^2 \,\forall\, i.$$

Suppose we know that $\mathrm{Var}\left[\varepsilon_i \mid \boldsymbol{x}_i\right] = \boldsymbol{x}_i'\boldsymbol{\beta}$. The test consists of the following steps. 1. Estimate $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$, and construct $\hat{\boldsymbol{\varepsilon}}$. 2. Square each term of the vector $\hat{\boldsymbol{\varepsilon}}$. Use the squared residuals as estimators of the elements on the diagonal of $\mathrm{Var}\left[\boldsymbol{\varepsilon} \mid \boldsymbol{X}\right]$. Treat this vector as the dependent variable, and $\boldsymbol{X}$ as the matrix of independent variables. Carry out OLS regression. 3. Compute $LM = nR^2$. The test statistic has an asymptotic χ^2 distribution with $J - 1$ degrees of freedom where J is the number of regressors including the column of ones.

FGLS estimator or the HCE?

If we detect heteroskedasticity, should we use the FGLS estimator or the HCE of the standard errors? In the FGLS approach, we first estimate the error covariance structure, and then use it to estimate the true coefficients. The estimation is on the transformed model. Hence, we alter the coefficient estimates. We also alter the standard errors since they depend on the altered coefficient estimates. The HCE does not do anything to the coefficient estimates. It acknowledges heteroskedasticity and accounts for it only in the standard errors. The advantage of the HCE is that we do not need to figure out the covariance structure. On the other hand, if it is possible to obtain a reasonable estimate of the error covariance structure, FGLS is better.