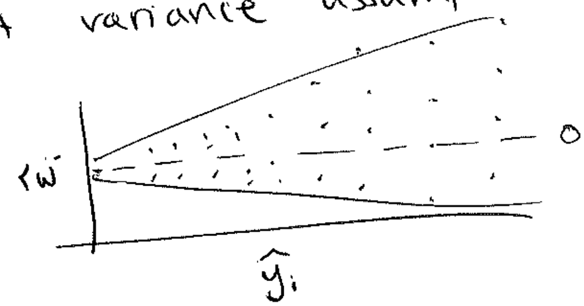
Generalized Least Squares

weighted least squares

So far, we have assumed

$$\underset{\sim}{y} \sim N_n(\underset{=}{X}\underset{\sim}{\beta}, \sigma^2 \underset{\equiv}{I})$$

However, after we fit a model we see the constant variance assumption violated:



← this is a case of <u>heteroskedacity</u> (non-constant variance)

Evidence that $var(\underset{\sim}{\varepsilon}) \neq \sigma^2 \underset{\equiv}{I}$.

Q: What should we do?

A: Change the <u>model</u>!

Instead, let's assume $var(\underset{\sim}{\varepsilon}) = \underset{\equiv}{\Sigma}$ ← need not be diagonal

<u>Examples</u>:

(1) $$\underset{\equiv}{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots \\ \rho & 1 & & & \\ \rho^2 & & \ddots & & \\ & & & & 1 \end{bmatrix}, \quad \rho \in [-1, 1]$$

<u>autocorrelation</u>
"AR-1" or "AR(1)"
- "autoregressive model of order 1"
- useful for time-correlated outcome

(2) $$\underset{\equiv}{\Sigma} = \begin{bmatrix} V_1 & & 0 \\ & \ddots & \\ 0 & & V_n \end{bmatrix} \quad V_i = f(x_i^T \beta)$$

← mean-variance relation.
ex: large expected value ⇒ large variance

(3) $$\underset{\equiv}{\Sigma} = \sigma^2 \begin{bmatrix} 1/m_1 & & 0 \\ & \ddots & \\ 0 & & 1/m_n \end{bmatrix}$$

diff. variances; diff observations have different levels of measurement precision.

ex: if my observations themselves are means of some other data I don't have, each w/ variable # obs.

If $\text{var}(\underline{\varepsilon}) = \underline{\underline{\Sigma}}$, then $\underline{y} \sim N_n(\underline{\underline{X}}\underline{\beta}, \underline{\underline{\Sigma}})$

Under this model, what is $\text{var}(\hat{\underline{\beta}}_{OLS})$?

Recall: $\hat{\underline{\beta}}_{OLS} = (\underline{\underline{X}}^T\underline{\underline{X}})^{-1}\underline{\underline{X}}^T \underline{y}$

$$\text{var}(\hat{\underline{\beta}}_{OLS}) = (\underline{\underline{X}}^T\underline{\underline{X}})^{-1}\underline{\underline{X}}^T \underbrace{\text{var}(\underline{y})}_{\underline{\underline{\Sigma}}}\underline{\underline{X}}(\underline{\underline{X}}^T\underline{\underline{X}})^{-1}$$

An aside:

If $\text{var}(\varepsilon_i) = \sigma^2$, then $\text{var}\left[(\sigma^2)^{-1/2}\varepsilon_i\right] = \text{var}\left(\frac{\varepsilon_i}{\sigma}\right) = \frac{1}{\sigma^2}\text{var}(\varepsilon_i) = 1$

Similarly, if $\text{var}(\underline{\varepsilon}) = \underline{\underline{\Sigma}}$, then

$$\text{var}(\underline{\underline{\Sigma}}^{-1/2}\underline{\varepsilon}) = \underline{\underline{\Sigma}}^{-1/2}\underline{\underline{\Sigma}}\,\underline{\underline{\Sigma}}^{-1/2} = \underline{\underline{I}}$$

Q: where does this come from?

A: when $\underline{\underline{\Sigma}}$ is symmetric, positive definite, then $\text{SVD}(\underline{\underline{\Sigma}}) = \underline{\underline{U}}\,\underline{\underline{D}}\,\underline{\underline{U}}^T$ where ($\underline{\underline{U}}$ orthogonal, $\underline{\underline{D}}$ diagonal).

then $\underline{\underline{\Sigma}}^{-1/2} = \underline{\underline{U}}\,\underline{\underline{D}}^{-1/2}\underline{\underline{U}}^T$

and it is easy to check that $\underline{\underline{\Sigma}}^{-1/2}\underline{\underline{\Sigma}}\,\underline{\underline{\Sigma}}^{-1/2} = \underline{\underline{I}}$.

We know $\hat{\underline{\beta}}_{OLS}$. What is $\hat{\underline{\beta}}_{GLS}$?

Computing $\hat{\beta}_{GLS}$ using `lm`:

Our model: $\underset{\sim}{y} \sim N_n(\underline{\underline{X}}\underline{\beta}, \underline{\underline{\Sigma}})$. Equivalently,

$y = \underline{\underline{X}}\underline{\beta} + \underline{\varepsilon}$ for $\underline{\varepsilon} \sim N_n(\underline{0}, \underline{\underline{\Sigma}})$.

"TRICK": multiply both sides by $\underline{\underline{\Sigma}}^{-1/2}$;

$$\underbrace{\underline{\underline{\Sigma}}^{-1/2} y}_{\tilde{y}} = \underbrace{\underline{\underline{\Sigma}}^{-1/2} \underline{\underline{X}}}_{\tilde{\underline{\underline{X}}}} \underline{\beta} + \underline{\underline{\Sigma}}^{-1/2} \underline{\varepsilon} \quad \& \quad var(\underline{\underline{\Sigma}}^{-1/2} \underline{\varepsilon}) = \underline{\underline{I}}$$

$\Rightarrow$ $\tilde{y} \sim N_n(\tilde{\underline{\underline{X}}}\underline{\beta}, \underline{\underline{I}})$

and the OLS estimate based on $\tilde{y}, \tilde{\underline{\underline{X}}}$ is

$$\begin{aligned}
\hat{\beta} &= (\tilde{\underline{\underline{X}}}^T \tilde{\underline{\underline{X}}})^{-1} \tilde{\underline{\underline{X}}}^T \tilde{y} \\
&= (\underline{\underline{X}}^T \underline{\underline{\Sigma}}^{-1/2} \underline{\underline{\Sigma}}^{-1/2} \underline{\underline{X}})^{-1} \underline{\underline{X}}^T \underline{\underline{\Sigma}}^{-1/2} \underline{\underline{\Sigma}}^{-1/2} y \\
&= \boxed{(\underline{\underline{X}}^T \underline{\underline{\Sigma}}^{-1} \underline{\underline{X}})^{-1} \underline{\underline{X}}^T \underline{\underline{\Sigma}}^{-1} y \equiv \hat{\beta}_{GLS}.}
\end{aligned}$$

Note:
- if $\underline{\underline{\Sigma}} = \underline{\underline{I}}\sigma^2$, then $\hat{\beta}_{GLS} = \hat{\beta}_{OLS}$
- if $\underline{\underline{\Sigma}} \neq \underline{\underline{I}}\sigma^2$, then $\hat{\beta}_{GLS} \neq \hat{\beta}_{OLS}$

(*) More generally, $var(\hat{\beta}_{GLS}) \leq Var(\hat{\beta}_{OLS})$

$\hat{\beta}_{GLS}$ is $\underset{\text{Best Linear Unbiased Estimator}}{\underbrace{\text{BLUE}}}$

Exercise

(i) Show $\mathbb{E}[\hat{\beta}_{GLS}] = \beta$

(ii) Show $var(\hat{\beta}_{GLS}) = (\underline{\underline{X}}^T \underline{\underline{\Sigma}}^{-1} \underline{\underline{X}})^{-1}$

What objective function is $\hat{\beta}_{GLS}$ minimizing?

$$\hat{\beta}_{GLS} = \underset{\beta}{\text{argmin}} \underbrace{(y - \underline{\underline{X}}\beta)^T \underline{\underline{\Sigma}}^{-1} (y - \underline{\underline{X}}\beta)}_{\text{weighted RSS}}$$

Note: same as $\underset{\beta}{\text{argmin}} (\tilde{y} - \tilde{\underline{\underline{X}}}\beta)^T (\tilde{y} - \tilde{\underline{\underline{X}}}\beta)$

Another view:

$y \sim N_n(\underline{\underline{X}}\beta, \underline{\underline{\Sigma}})$, so the density of y is

$$p(y) = (2\pi)^{-n/2} \det(\underline{\underline{\Sigma}})^{-1/2} \exp\left\{ -\frac{1}{2}(y - \underline{\underline{X}}\beta)^T \underline{\underline{\Sigma}}^{-1} (y - \underline{\underline{X}}\beta) \right\}$$

The log-likelihood can be written:

$$\log L(\beta, \underline{\underline{\Sigma}}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log\det(\underline{\underline{\Sigma}}) - \frac{1}{2}(y - \underline{\underline{X}}\beta)^T \underline{\underline{\Sigma}}^{-1} (y - \underline{\underline{X}}\beta)$$

unknown parameters

to maximize log-likelihood of β, I take derivative & set $= 0$:

$$\frac{d}{d\beta} \log L(\beta, \underline{\underline{\Sigma}})$$

---

but notice! $\log L(\beta; \underline{\underline{\Sigma}})$ (viewed as a function of β) is equal to

$$\underbrace{-\frac{1}{2}(y - \underline{\underline{X}}\beta)^T \underline{\underline{\Sigma}}^{-1} (y - \underline{\underline{X}}\beta)}_{-\text{ weighted RSS}} + \text{some irrelevant constant.}$$

maximizing negative of a function is same as minimizing the function