Robust SEs

For the full picture, we'll have to wait until LT (and matrices!). This is just a taste.

- I've mentioned in class that when we get to LT, we're going to show <u>mathematically</u> when you should use each of the tools you'll learn in MT. We do this by saying "under the assumptions of X, Y, Z..., (for example) OLS has properties A, B, C...". Then, based on the assumptions you think are true, you know which good (or bad!) properties each method has, so you can choose the best one.

- One important assumption: $\underbrace{\text{homo}}_{\text{"same"}}\underbrace{\text{scedasticity}}_{\text{"related to variance"}}$

  another: no $\underbrace{\text{auto}}_{\text{"with itself"}}\ \underbrace{\text{correlation}}_{\text{"correlation" (duh.)}}$

- <u>Formally</u>: If $y_i = \alpha + x_{1i}\beta_1 + \dots + x_{ki}\beta_k + \varepsilon_i$

  Homoscedasticity: $Var(\varepsilon_i)$ does <u>not</u> depend on $\{x_{1i} \dots x_{ki}\}$

  No autocorrelation: $\varepsilon_i, \varepsilon_j$ are uncorrelated for $i \neq j$

- (Ignore for now, but in case you look at these after some LT material)

If we look at the covariance matrix $\mathbb{E}(\varepsilon\varepsilon')$, then homoscedasticity means there is a constant (σ^2) on the diagonal. No autocorrelation means it's 0 everywhere else. Between the two assumptions, this means

$$\mathbb{E}(\varepsilon\varepsilon') = \begin{bmatrix} \sigma^2 & 0 & 0 & & 0 \\ 0 & \sigma^2 & 0 & \ddots & \\ 0 & 0 & \sigma^2 & & 0 \\ \vdots & & & \ddots & \\ & 0 & & & \sigma^2 \end{bmatrix} = \underset{\substack{\uparrow \\ \text{Identity} \\ \text{matrix}}}{\sigma^2 I}$$

As the diagonal terms $[\mathbb{E}(\varepsilon\varepsilon')]_{ii}$ are $Var(\varepsilon_i)$ and the off-diagonal terms $[\mathbb{E}(\varepsilon\varepsilon')]_{ij}$, $i \neq j$ are the covariances $cov(\varepsilon_i, \varepsilon_j)$

- The idea estimating OLS with Robust SEs is a different tool to doing OLS with usual SEs.

  ↳ So, when is each one better?

  ↳ short answer: if we have homoscedasticity and no autocorrelation, then use usual SEs. otherwise use Robust SEs.

- Note:
  - The opposite of homoscedastic is heteroscedastic.
    "different" "related to variance"
    The opposite of no autocorrelation is (of course!) just "autocorrelation"
  - In stata if our dependent / outcome / LHS variable is y and the regressor / covariate / independent / RHS variable is x, then:

OLS usual SEs: reg y x

OLS robust SEs: reg y x, robust

'Robust' in stata means the ones which deal with heteroscedasticity, but still presume no AC

note this is just an option we add to the usual command

  - There are many other methods which provide a halfway-house between usual SEs and robust SEs. For example, cluster SEs are when we think some people have $\mathrm{cor}(\varepsilon_i, \varepsilon_j) \neq 0$, but others have $\mathrm{cor}(\varepsilon_i, \varepsilon_k) = 0$

This is useful if, for example, we think some people might have similar outcomes because of their location. In development economics, for example, we often "cluster" for people from the same village; maybe what happens to people from the same village is correlated with each other, because they interact with each other!

(ignore until LT). The assumptions for each are best expressed through the covariance matrix:

"Robust"

Homoscedastic:

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

All σ^2 same

Heteroskedastic, but no AC:

$$\begin{bmatrix} \sigma_1^2 & 0 & & 0 \\ 0 & \sigma_2^2 & & \\ & & \ddots & 0 \\ 0 & & 0 & \sigma_n^2 \end{bmatrix}$$

σ_i^2 different potentially

Cluster:

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & 0 & 0 & 0 & 0 \\ \sigma_{21} & \sigma_2^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & \sigma_{34} & \sigma_{35} & 0 \\ 0 & 0 & \sigma_{43} & \sigma_4^2 & \sigma_{45} & 0 \\ 0 & 0 & \sigma_{53} & \sigma_{54} & \sigma_5^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_6^2 \end{bmatrix}$$

Fully Robust

$$\begin{bmatrix} \text{ANYTHING} \end{bmatrix}$$

There are others too

- <u>Question</u>: why so many types?

↳ There's a fundamental trade-off (as we'll see in LT):-

<u>In general</u>, the more you assume

- The <u>better</u> you estimate if your assumptions are <u>correct</u>
- The <u>worse</u> you do if your assumptions are <u>wrong</u>

↳ "Better" and "Worse" show up in all sorts of ways, as we will see in the LT: if you assume the wrong things you may lose some of the nice properties an estimator has; if you assume too little your estimates $\hat{\beta}$ might be very inaccurate (eg very high variance → very likely to be far from the truth)

↳ The <u>art</u> of applied econometrics is getting this balance correct. We'll talk much more about it in LT!

More advanced: How do robust SE work?

For full details we need matrices. However I can do a simpler example without them. Suppose:

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

no intercept α, only one regressor x_i

Then we can write the OLS estimator $\hat{\beta}$ as:

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i^2 \beta + \sum x_i \varepsilon_i}{\sum x_i^2}$$

$$= \beta + \frac{\sum x_i \varepsilon_i}{\sum x_i^2}$$

As adding a constant (β) doesn't change the variance:

$$Var(\hat{\beta}) = Var\left(\frac{\sum x_i \varepsilon_i}{\sum x_i^2}\right)$$

$$= \left[\frac{1}{\sum x_i^2}\right]^2 Var\left(\sum x_i \varepsilon_i\right)$$

Supposing the x_i are non-random: we'll see more about this assumption in LT

$$= \left[\frac{1}{\sum x_i^2}\right]^2 \sum x_i^2 \, Var(\varepsilon_i)$$

No AC

under homoscedasticity

$$Var(\varepsilon_i) = \sigma^2 \quad \forall i$$

Therefore:

$$Var(\hat{\beta}) = \left[\frac{1}{\sum x_i^2}\right]^2 \sum x_i^2 \sigma^2$$

$$= \sigma^2 \cdot \left[\frac{1}{\sum x_i^2}\right]$$

It turns out that $\hat{\sigma}^2 = \frac{1}{n} \sum (y_i - x_i \hat{\beta})^2$ is a "good" estimator for σ^2. So for usual SEs, stata estimates $var(\hat{\beta})$ by $\hat{\sigma}^2 \cdot \left(\frac{1}{\sum x_i^2}\right)$

However under heteroscedasticity and/or autocorrelation we can't do that last cancellation step. If we assume no AC, but allow for heteroscedasticity, the most we can cancel to is:

$$Var(\hat{\beta}) = \left[\frac{1}{\sum x_i^2}\right]^2 \sum x_i^2 \, Var(\varepsilon_i)$$

Stata's Robust SE (known as 'white SE', after the inventor), estimate this directly

We can see immediately two things:

- Robust SEs are "good" even under homoscedasticity and no AC.
- **But** if we **do** have homoscedasticity and no AC, then estimating using usual SEs requires us to estimate the 'simple' object σ^2 rather than a 'complex' object like the un-cancelled form; so it does a better job.

↳ comes back to the same idea: choose the **right** assumptions

So why always run reg y x, robust?

Because the assumption of homoscedasticity is **very** unrealistic! For many other methods there's a genuine tradeoff, but for robust vs. usual SEs, robust is **almost always** the **right** choice! But as you do more Metrics, you'll realise the other methods (cluster, fully robust, etc) all have right times and places to use them in!