HETEROSKEDASTICITY

The problem: $Cov(\epsilon) = \Omega$,
where Ω is a diagonal matrix
$[Cov(\epsilon_t, \epsilon_{t-i}) = 0 \quad \forall \; i \neq 0]$
but the elements on diagonal are $\neq$.

i.e. violates assumption of constant variance

So now assume: $\epsilon_t \sim N(0, \sigma_t^2)$

- are independent
- but σ_t^2 depends on t

i.e.
$$\Omega = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3^2 & & 0 \\ & \vdots & & & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_T^2 \end{bmatrix}$$

* This typically arises in cross-sectional data when the data ~~varys~~ vary a great deal.
In this case, the variances accross the data will not likely be equal.

e.g. $Var(\epsilon_t)$ may depend on the scale.
→ may be correlated with X_t ...
→ [illegible]

In this case — let $y = \alpha + \beta x + \epsilon$

<u>Note:</u> If $\exists$ correlation between x & ϵ, that's serious! —(more later)

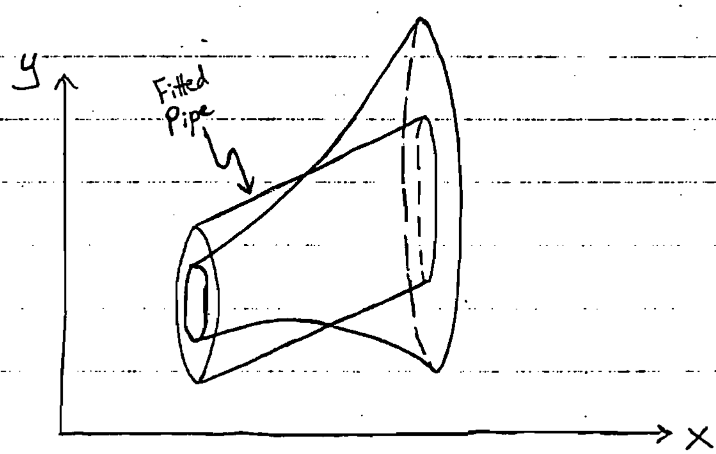
If $\exists$ correlation between x & <u>$Var(\epsilon)$</u>, we can do something.

<u>Picture:</u>

(1)

- diameter of pipe varies with x..

(observations fall within (cone) pipe)



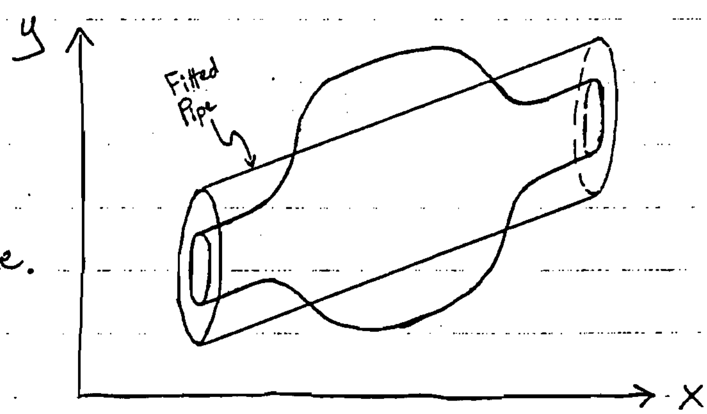
(2)

If problem ignored, <u>OLS</u> fits best line,

& estimates an <u>average</u>, <u>constant</u> variance.

$\left(s^2 = \frac{1}{T-K} SSE \right)$

= avg diameter of pipe.



Consequences: (i) OLS $\hat{\beta}$ unbiased, consistent, inefficient.

(ii) s^2 biased & inconsistent. (obvious from pictures)

(iii) t & F statistics wrong.

With Picture 1,

s^2 = avg. diameter
gives false impression that
data is more confined. (than it really is)

For large X, s^2 is too small.

* ∴ Make statements about precision (t-ratios)
which are overstated,
- not conservative,
- wrong!

With Picture 2,

data is apparently more confined
(to the narrower pipe)
than s^2 indicates.

∴ s^2 too big,
statements are too conservative.

☆ If we must err,
better to err on conservative side!

Solution: GLS

We need to find Ω, Ω^{-1},
and the transformation, V, $\ni$ $V'V = \Omega^{-1}$.

$$\text{Again, } Cov(\epsilon) = \Omega = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & 0 \\ & \vdots & & \\ 0 & 0 & \cdots & \sigma_T^2 \end{bmatrix}$$

* [for simplicity, include σ^2 in Ω (formerly, had $Cov(\epsilon) = \sigma^2 \Omega$).]

$$\Longrightarrow \Omega^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2^2} & & 0 \\ & \vdots & & \\ 0 & 0 & \cdots & \frac{1}{\sigma_T^2} \end{bmatrix}$$

$$\Longrightarrow \Omega^{-1} = V'V \quad \text{where} \quad V = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & & 0 \\ & \vdots & & \\ 0 & 0 & \cdots & \frac{1}{\sigma_T} \end{bmatrix}$$

We have:

$Y = X\beta + \epsilon$; $\quad y_t = \beta_1 X_{t1} + \beta_2 X_{t2} + \cdots + \beta_K X_{tK} + \epsilon_t$

The GLS transformation:

$VY = (VX)\beta + V\epsilon$; $\quad \frac{y_t}{\sigma_t} = \beta_1\left(\frac{X_{t1}}{\sigma_t}\right) + \beta_2\left(\frac{X_{t2}}{\sigma_t}\right) + \cdots + \beta_K\left(\frac{X_{tK}}{\sigma_t}\right) + \left(\frac{\epsilon_t}{\sigma_t}\right)$

$t = 1, \ldots, T$

$$\Rightarrow \quad Var\left(\frac{\epsilon_t}{\sigma_t}\right) = \frac{1}{\sigma_t^2} Var(\epsilon_t) = \frac{1}{\sigma_t^2}\sigma_t^2 = 1$$

Then simply apply OLS to the transformed system.

This is a special case of GLS called "<u>Weighted Least Squares</u>."

<u>What if Ω is unknown?</u> — the usual case!
(i.e. if σ_t^2 is unknown $\forall t$)

$\exists$ no simple, well-accepted assumption about the σ_t^2 values. — [as there is for autocorr

— Look at the data.
— Look at the residuals of OLS.
— Appropriate assumption will depend on particular problem.

Possible Assumptions for Heteroskedasticity

Example 1: Assume $y_t = \alpha + \beta x_t + \epsilon_t$
with $Var(\epsilon_t) = \sigma^2 x_t^2$.

[i.e. $Var(\epsilon_t)$ is proportional to x_t^2]

Then regress

$$\left(\frac{y_t}{x_t}\right) = \alpha\left(\frac{1}{x_t}\right) + \beta + \left(\frac{\epsilon_t}{x_t}\right)$$

e.g. in Gini Ratio computer problem

$$\Longrightarrow Var\left(\frac{\epsilon_t}{x_t}\right) = \frac{1}{x_t^2} Var(\epsilon_t) = \frac{1}{x_t^2}(\sigma^2 x_t^2)$$

Note: $\exists$ problems with this if you have more than one regressor.
In that case, pick one x_{ti};
then other transformed regressors become $\frac{x_{t1}}{x_{ti}}, \frac{x_{t2}}{x_{ti}}, \ldots$
— still shaky.

Example 2: $Y = X\beta + \epsilon$; t^{th} obs. -- $\underset{(1\times K)(K\times 1)}{y_t = x_t \beta + \epsilon_t}$

Assume
$$Var(\epsilon_t) = \sigma^2[E(y_t)]^2 = \sigma^2(x_t\beta)^2$$

[i.e. $Var(\epsilon_t)$ is proportional to $(x_t\beta$

Here you aren't picking just one x_{ti}.

If $(X_t\beta)$ is known, regress

$$\left(\frac{y_t}{X_t\beta}\right) = \beta_1\left(\frac{X_{t1}}{X_t\beta}\right) + \cdots + \beta_K\left(\frac{X_{tK}}{X_t\beta}\right) + \left(\frac{\epsilon_t}{X_t\beta}\right)$$

$$\longrightarrow Var\left(\frac{\epsilon_t}{X_t\beta}\right) = \frac{1}{(X_t\beta)^2}(X_t\beta)^2\sigma^2 = \sigma^2.$$

☆ However, β is unknown!
That's why we're running regression!

<u>To apply GLS</u>;

(i) Run OLS on original specification,
get $\hat{\beta}$ — unbiased, consistent, but inefficient.

(fitted values, $\hat{y}_t$)

(ii) Use $(X_t\hat{\beta})$ as weights → GLS transformation.
get $\tilde{\beta}$ — efficient relative to $\hat{\beta}$.

☆ Since $\hat{\beta}$ is consistent,
$\tilde{\beta}$ will be <u>asymptotically</u> <u>efficient</u>.
(has same asymptotic distribution as
GLS with Ω known --- the top regression above)

This can be an iterative procedure,
but each stage has same asymptotic dist.

Amemiya, <u>JASA</u>, 1973
shows GLS inefficient relative to <u>MLE</u>.
This is inefficient procedure since you're using an estimate of β to estimate β! ∴ says grind out MLE. — more work, bu reasonable altern

Example 3: Grouped Data – by states

let $y_{ij} = \alpha + \beta x_{ij} + \epsilon_{ij}$; $\underbrace{\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)}_{\text{satisfies ideal cond.}}$

[observation for j^{th} person in i^{th} state]

$i = 1, \ldots, 50$ ← represents state

$j = 1, \ldots, n_i$ ← n_i represents # of observations in i^{th} state.

Suppose we don't observe y_{ij}, but rather observe $\bar{y}_i$

where $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$ = avg of obs. in state i

$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$ = aggregated avg. data over people in state i

☆ i.e. Suppose data are aggregated.

Then the model we are forced to work with is

$$\bar{y}_i = \beta \bar{x}_i + \bar{\epsilon}_i \quad \text{where } \bar{\epsilon}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \epsilon_{ij}$$

Note: The parameters, β, in the aggregated relationship are in the same functional form as in the disaggregated relationship, since the aggregation procedure is a linear transformation.

Before, $\epsilon_{ij} \overset{\boxed{iid}}{\sim} N(0, \sigma^2)$

Now, $\bar{\epsilon}_i \sim N(0, \frac{\sigma^2}{n_i})$

$\longrightarrow$ $\exists$ heteroskedasticity, since $n_i \neq n_j \; \forall \; i \neq j$.

* Larger (more populous) States (more obs.) will have a smaller variance of the avg. disturbance.

How to solve this? GLS!

Find a V!

Need a transformation that will make $\bar{\epsilon}_i$ have the same Variance $\forall$ state.

$Var(\bar{\epsilon}_i) = \frac{\sigma^2}{n_i}$

Consider $\sqrt{n_i}\,\epsilon_i$; $Var(\sqrt{n_i}\,\epsilon_i) = n_i \frac{\sigma^2}{n_i} = \sigma^2$!

Thus ; $VY = (VX)\beta + V\epsilon$

$\Longrightarrow (\sqrt{n_i}\,\bar{y}_i) = \sqrt{n_i}\,\alpha + (\sqrt{n_i}\,\bar{x}_i)\beta + (\sqrt{n_i}\,\bar{\epsilon}_i)$

Run OLS on this $\longrightarrow \tilde{\beta}$.

* Relevant problem with aggregated data!

See Johnston, Maddala, Goldfield & Quandt - <u>Nonlinear Methods in Econometrics</u>.

Look at <u>Residuals</u>!

OLS $\hat{\beta}$ is unbiased, consistent.

Picture 1 or Picture 2 ?

Which kind of assumption is appropriate?

Example 1, 2, 3, ...

Is $Var(\epsilon_t)$ correlated with something ?

<u>Tests:</u>

Ask: Should you Regress Residuals on X_t ?

$$e_t = \alpha + \beta X_t + u_t$$

No! X_t & ϵ_t <u>un</u>correlated, or serious problems!!

(i) Regress <u>square</u> of OLS Residuals on X_t.

$$e_t^2 = \alpha + \beta X_t + u_t \qquad \text{(significant } F\text{ ?)}$$

→ $Var(\epsilon_t)$ = <u>second</u> moment = $E(\epsilon_t^2)$;

Does <u>this</u> depend on X_t ?

$\hat{\beta} > 0 \Rightarrow$ Pict... 1 [illegible]

(ii) Regress $e_t^2 = \alpha + \beta(X_t^2) + u_t$ (significant F?)

$\beta \rightarrow$ $Var(\epsilon_t)$ proportional to (X_t^2).

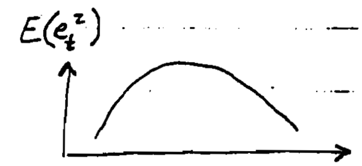
If $\hat{\beta} > 0 \Rightarrow$ Picture ① again.

[Here the transformation in Example 1 is appropriate.]

(iii) Regress $e_t^2 = \alpha + \beta_1 X_t + \beta_2 X_t^2 + u_t$

If $\beta_2 < 0$, $\Longrightarrow$ Picture ②.

i.e. Variance first ↑es, then ↓es.



Again, OLS on this kind of heteroskedasticity <u>overstates</u> s^2 & ∴ gives overly conservative tests. — Can correct this with GLS.

(popular test)

(iv) <u>Goldfield & Quandt Test</u>, [<u>JASA</u>, 1965]

Split the sample, do OLS on each $\Rightarrow s_1^2$ & s_2^2.

$$Y_1 = X_1\beta_1 + \epsilon_1 \quad (T_1 \text{ obs.})$$
$$Y_2 = X_2\beta_2 + \epsilon_2 \quad (T_2 \text{ obs.})$$

Then $\frac{s_1^2}{s_2^2} \sim F_{T_1-K,\, T_2-K}$ under $H_0: \sigma_1^2 = \sigma_2^2$.