Heteroscedasticity.

CLRM asn — variance of the disturbance term in each obs is constant.

What does this mean?

Disturbance term in each obs has only one value so variance???

This variance refers to the potential distribution of the disturbance term before the sample is generated.

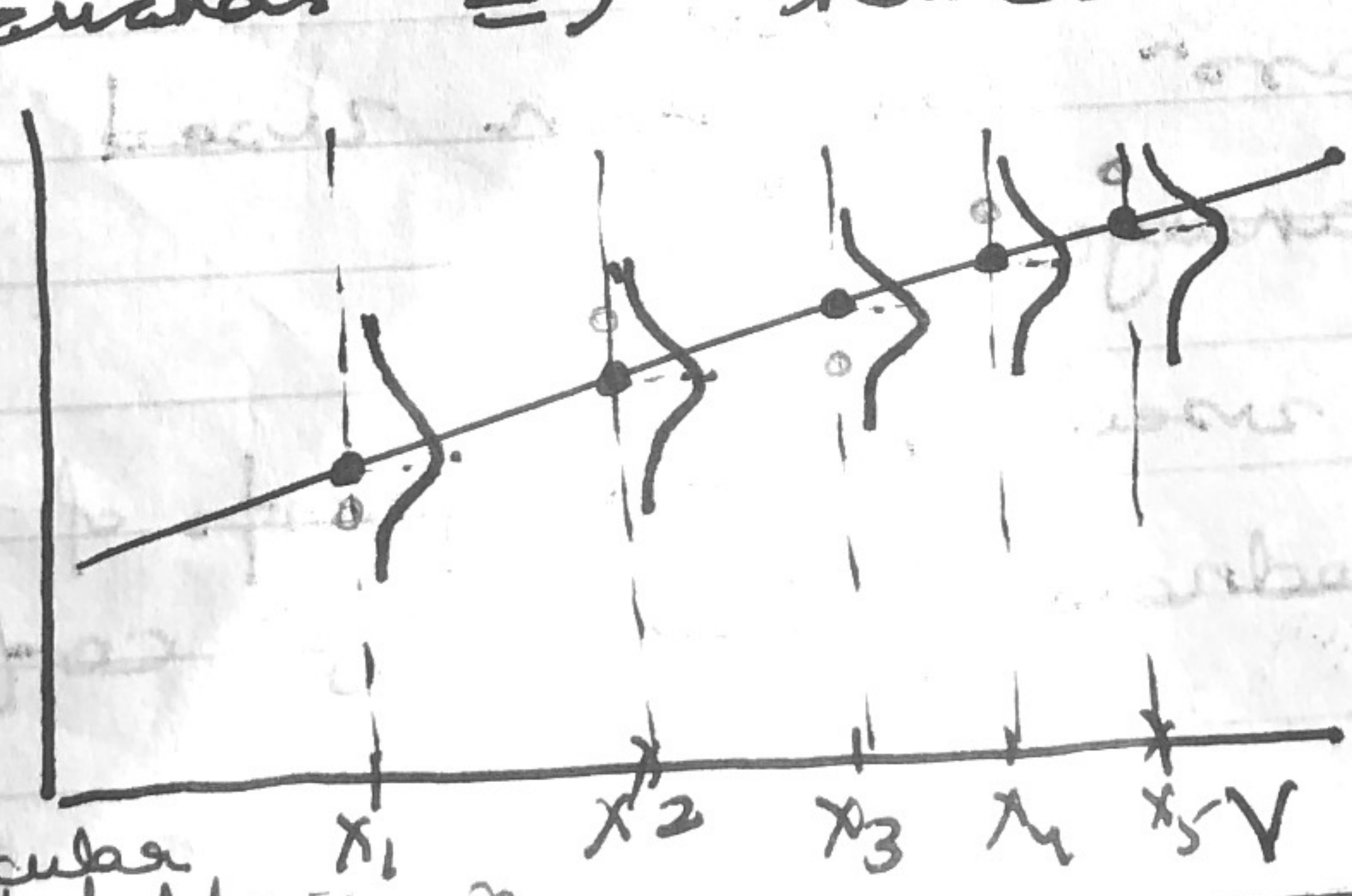
$$Y = B_1 + B_2 X + u_i$$

$$u_i \sim N(0, \sigma^2)$$

$\Rightarrow u_1, u_2, u_3, \ldots u_n$ are n observations drawn from prob^y distributions that have zero mean + same variance

The actual values of the disturbance term in the sample will be sometimes +ve, sometimes −ve, but there is no reason to anticipate a particularly erratic value in any given observation i.e prob^y of u reaching a +ve or negative value will be the same in all observations $\Rightarrow$ homoscedasticity

i.e the probability of getting a large +ve error should not be higher for any particular subpopulation

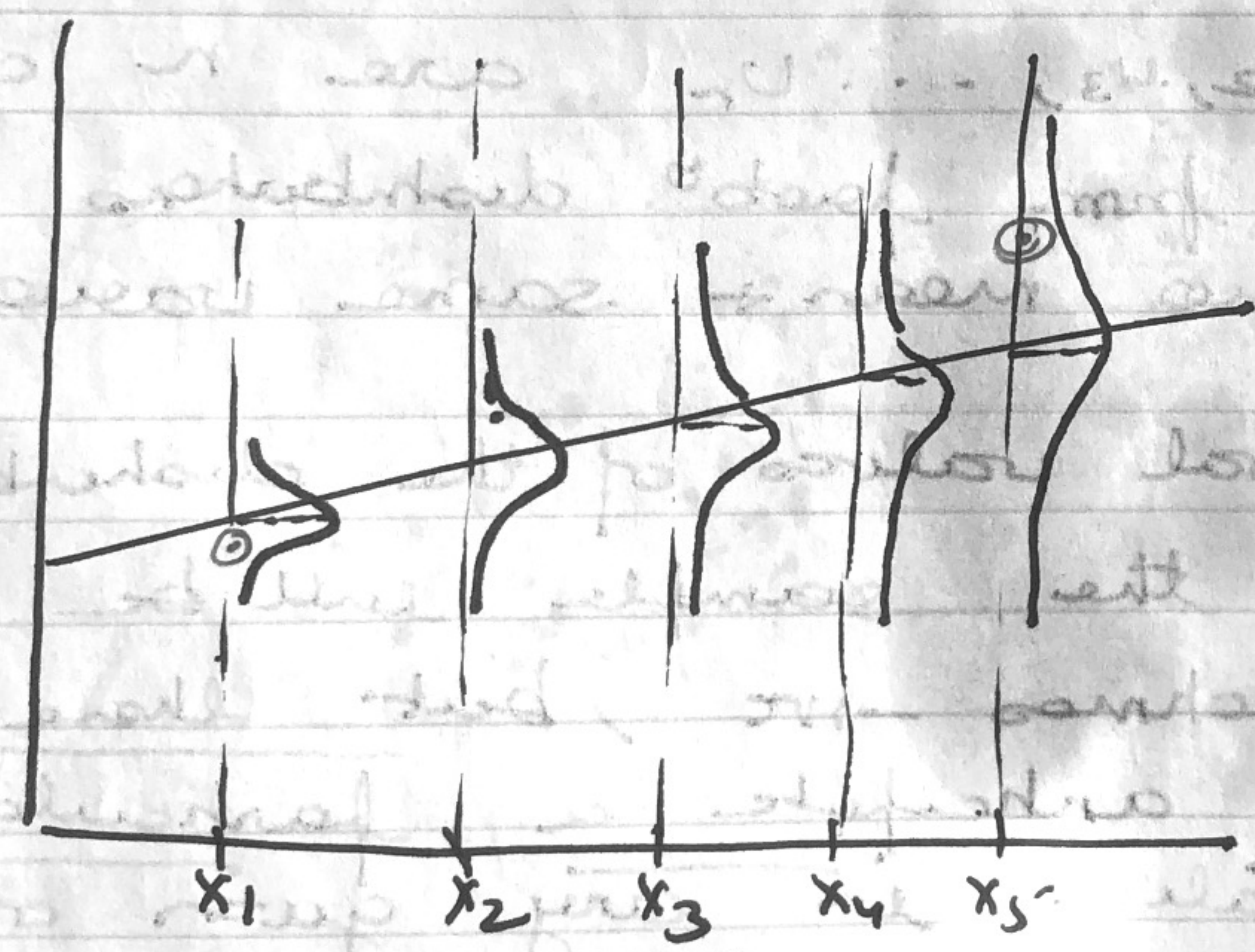


In some contexts may be reasonable to suppose that the potential distribution of the disturbance term is different for diff observations in the sample.

Below → potential distribution of the disturbance term is shows variance is ↑ as x ↑s.

This does not mean that the disturbance term will necessarily have a particularly large (+ive or negative) value in an obs where x is large, but that the prob of having an erratic value will be relatively high

⇒ heteroscedasticity.



Why does heteroscedasticity matter?

This assⁿ has not been used anywhere in the analysis

→ not used in the proofs of the unbiasedness of OLS regⁿ coeffs

1) We want the variances of the regn coeff to be as small as possible so that there is max precision
In the presence of homoscedasticity & other CLRM ass^ns OLS estimators are BLUE
but with heteroscedasticity, OLS est are inefficient because in principle, we can find other estimators that have smaller variances & still unbiased

2) Estimators of the std errors of the regn coeff will be wrong. They are computed on the ass^n that there is homoscedasticity. If this is not the case, the std errors are biased
⟹ t tests & F tests are invalid.
Quite likely that std errors are underestimated
⟹ t stat are overestimated
⟹ will end up getting significant tests when that may not be the case

Consequences of H.S

1. OLS still linear
2. " " unbiased
3. But they are inefficient (even in large samples)
4. Usual formulas for var for OLS estimator

are biased. Cannot tell the sign of the bias tho more likely to be a negative bias i.e OLS underestimates true variance of estimators

5. The bias arises because $\hat{\sigma}^2 = \frac{\Sigma e_i^2}{df}$, the estimator of σ^2 is ~~not~~ no longer an unbiased est of σ^2

6 Usual confidence intervals & hypo tests based on t & F dist are unreliable

$\Rightarrow$ cd lead to wrong conclusion

<u>Intuitive expl^n</u> <u>for</u> <u>inefficiency</u> of <u>OLS estimators</u>

An obs where the potential dist^n of disturbance term has a small std deviation will tend to lie close to the line $Y = B_1 + B_2 X$ & hence will be a good guide to the location of the line. By contrast an obs where the potential distribution has a large std deviation will be an unreliable guide to the location of the line

Some obs. (lower variance) are more informative than others (higher var)

OLS does not discriminate b/w the quality of the observations giving equal weight to each.
If we can find a way to give more weight to high quality obs & less to the unreliable ones we are likely to obtain a better fit.
=> more efficient estimation.

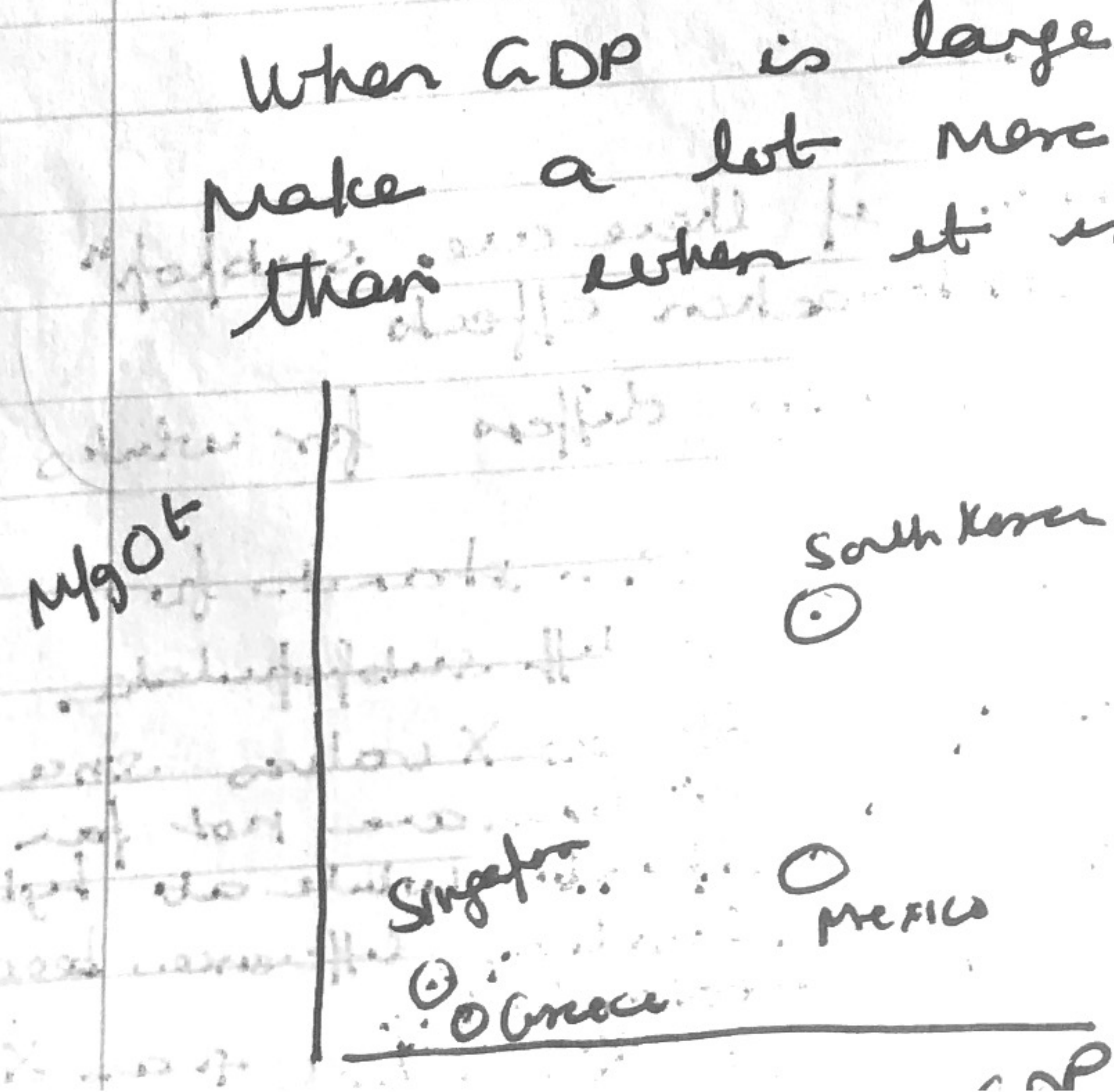
Possible causes of H.S

H.S will be a prob when the values of the variables in the sample (Y values) vary substantially in diff observations.

Suppose reg Value added in M/g on GDP using cross county data.
M/g O/t tends to be abt 15 to 25% of GDP variations being caused by comparative advantage & historical eco devt.

When GDP is large a 1% diff will make a lot more diff in absolute term than when it is small

Scale effects

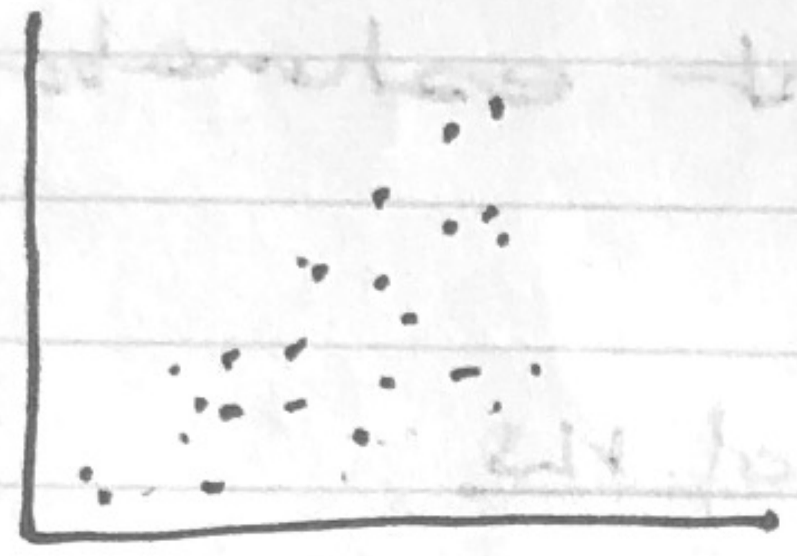


Singapore & Greece have relatively large & small m/g O/t but because both ones are small their variations from trend are small

Causes of H.S

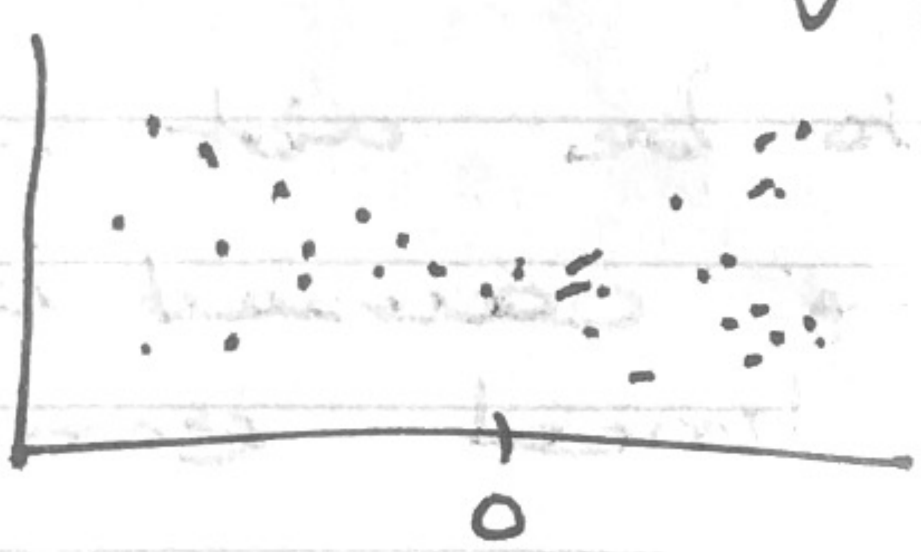
1. Scale effect – errors may ↑ as the value of an IV increases

eg annual family exp on vacations regressed upon annual family income



eg ~~share~~ Mfg output on GDP

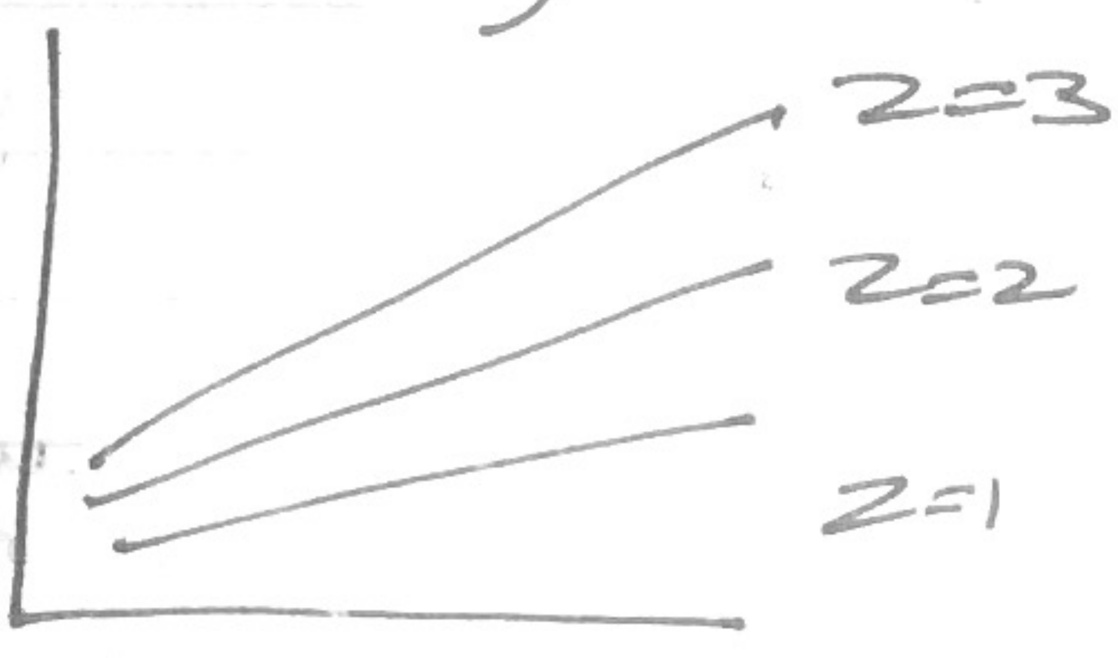
2. errors may also ↑ as values of an IV become more extreme in either direction
eg attitudes that range from extremely negative to extremely positive



– hourglass.

3. Measurement error

4. H.S can also occur if there are subpop^n differences or other interaction effects
(eg effect of income on exp differs for whites + Males)



for eg small medium + large firms sampled together

Z – stands for 3 diff subpopulations
At low X values since reg^n lines are not far apart while at high X values differences become substantial
⟹ spread of errors will ↑ as X ↑

5. Model misspecification can produce H.S
6. outliers

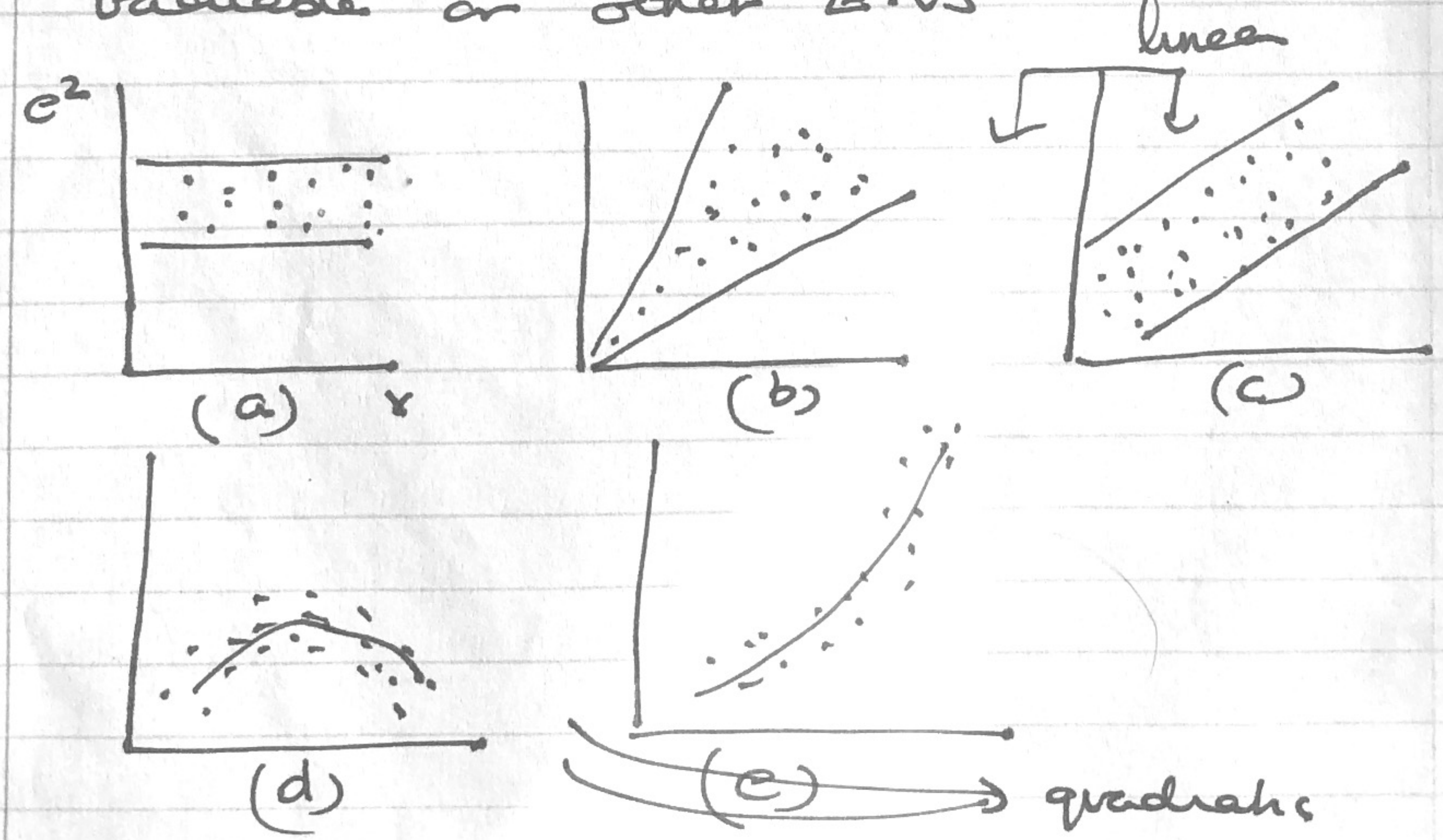
Detection of H.S

Detection is difficult because σ_i^2 can be known only if we have the entire Y popⁿ corresponding to the given X's

- typically, we have one Y value for a chosen X ⟹ we cannot determine the distⁿ of Y given X.

Diagnostic tools

1. Graphical Examination of Residuals
Use OLS to fit a line + obtain residuals
Plot the residuals against the dependant variable or other E.Vs



a – no pattern

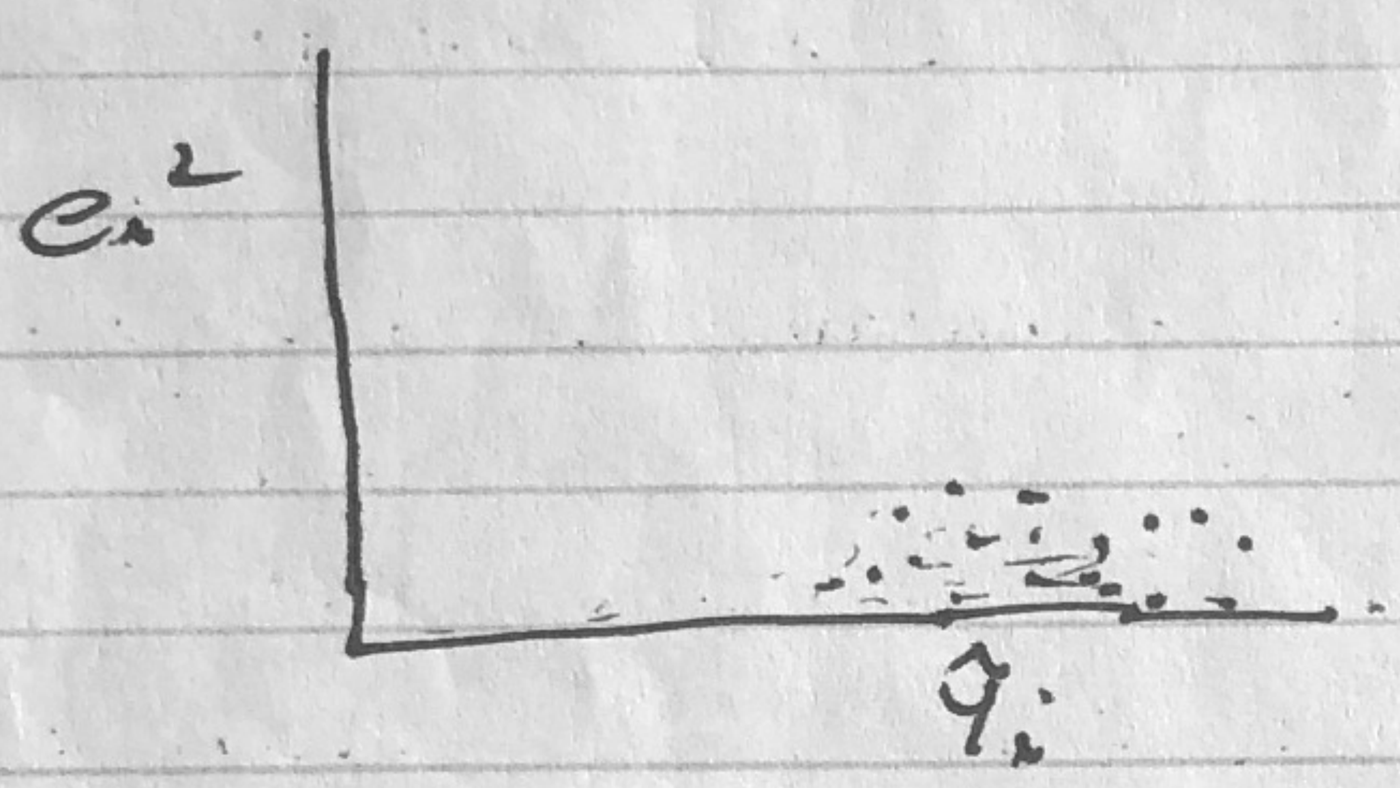
b to e – systematic relationship

So if squared residuals exhibit any one of the pattern from (b) to (e), HS is a possibility

But what if our regression is a multiple regression involving four X variables

$\Longrightarrow$ cd plot e_i^2 against each X

or plot e_i^2 against $\hat{Y}_i$
as $\hat{Y}_i$ is a linear comb^n of all X_i's



$\Longrightarrow$ indicates HS.

So now what?

<u>Park Test</u>

If HS is present $\Longrightarrow$ the heteroscedastic variance may be systematically related to one or more explanatory variables
To see if this is so, reg σ_i^2 on the X

$$\ln \sigma_i^2 = B_1 + B_2 \ln X_i + V_i.$$

But we don't know σ_i^2
Use e_i's as proxies

$$\ln e_i^2 = B_1 + B_2 \ln X_i + V_i.$$

May give up logs if some X values are negative

Steps in Park Test:

1. Run OLS reg^n despite HS
2 Obtain residuals e_i, square them + take their logs.

3. Run the reg^n

$$\ln e_i^2 = B_1 + B_2 \ln X_i + V_i$$

If more than one X_i can run this reg^n for each X_i or use $\hat{Y}_i$

4 Test $H_0: B_2 = 0 \Rightarrow$ no HS

- If reject H_0 then $\Rightarrow$ HS $\Rightarrow$ take remedial action
If cannot reject H_0 then
B_1 = value of common or homoscedastic variance

Limitation: The error term V_i itself may be heteroscedastic in which case we are back to square one

Gleijser Test

$$|e_i| = B_1 + B_2 X_i + V_i$$
$$|e_i| = B_1 + B_2 \sqrt{X_i} + V_i$$
$$|e_i| = B_1 + B_2 \frac{1}{X_i} + V_i$$

- suffers from same limitation
may be used in large samples

White's general H.S Test

H.S does not affect R^2 or adjusted R^2 estimate (since these estimate popn variances which are not conditional on X)

Consider the model

$$Y_i = B_1 + B_2 X_{2i} + B_3 X_{3i} + u_i.$$

whether or not there is HS the estd coeff are unbiased => R^2 is same

1. Use OLS to estimate the regn & get e_i's

2. Now run the auxillary regression

$$e_i^2 = A_1 + A_2 X_{2i} + A_3 X_{3i} + A_4 X_{2i}^2 + A_5 X_{3i}^2 + A_6 X_{2i} X_{3i} + v_i.$$

[add^al powers of X may also be included]

& since this test is based on R^2 => no prob

3. Obtain the R^2 of the above regn
Under the null hypo that there is no H.S (i.e all slope coeff are zero), White has shown that:

$$nR^2 \sim \chi^2_{k-1}$$

but t tests based on variance calculated in the presence of HS is biased => can't use t test

4. If the chisquare obtained exceeds the critical chi-square value at the chosen level of sig or if p value is low we can reject the H_0 & conclude that HS is present

Other tests

1. Spearman's rank correlation test
2. Goldfeld Quandt test
3. Bartlett's homogeneity of variance test
4. Peak test
5. Breusch Pagan test
6. CUSUMSQ Test

Remedial Measures:

We need to transform the model so that there is homoscedasticity
— What kind of transformation depends on whether the true error variance σ_i^2 is known or unknown.

When σ_i^2 is known: (WLS)

$$Y_i = B_1 + B_2 X_i + u_i.$$

Y_i ↓ hrly wages

X_i ↓ education (years of schooling)

Suppose σ_i is known

$$\frac{Y_i}{\sigma_i} = B_1\left(\frac{1}{\sigma_i}\right) + B_2\left(\frac{X_i}{\sigma_i}\right) + \left(\frac{u_i}{\sigma_i}\right)$$

Deflated both sides by σ_i

Let $\frac{u_i}{\sigma_i} = V_i.$

$$\text{Var } V_i = \frac{1}{\sigma_i^2} V(u_i) = \frac{\sigma_i^2}{\sigma_i^2} = 1$$

$\Rightarrow$ transformed model has a homo-scedastic error term

$\Rightarrow$ no HS

$\Rightarrow$ use OLS

~~Suppose~~ ~~[illegible]~~ ~~some variable~~

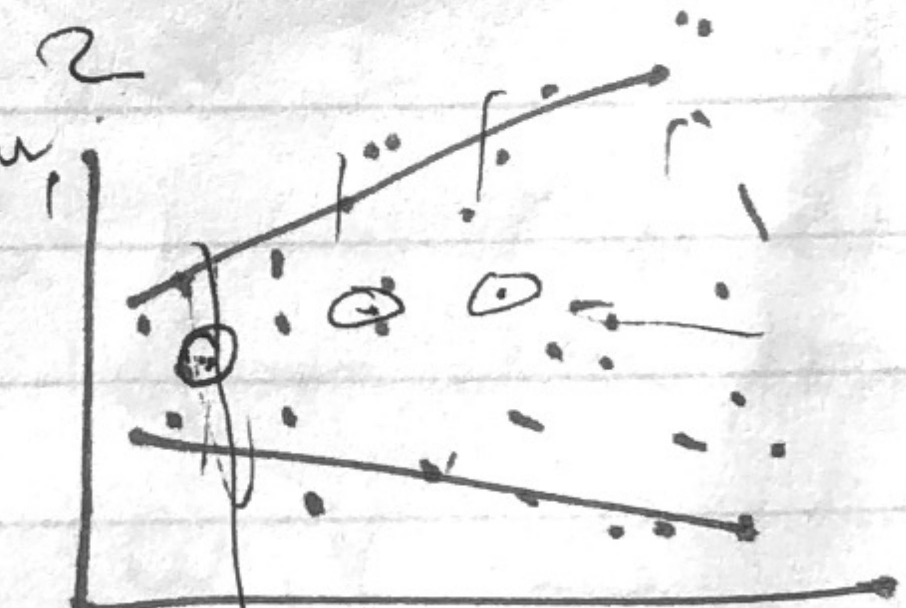
The OLS est^{on} of B_1 & B_2 are called WLS estimators since each Y & X observation is weighted by its own heteroscedastic std deviation σ_i.

<u>When true σ_i^2 is unknown.</u>

Rarely know σ_i^2, so will have to make some ass^n abt σ_i^2 & then transform the model. & then apply WLS.
(WLS is simply OLS applied to transformed data)

What ass^n's can we make abt the unknown error variance *

<u>Case I</u> : Error Variance is prop^{al} to X_i:
Square root transf^n.



Suppose the OLS residuals plotted against X show the shown pattern

$\Rightarrow$ error variance is prop^{al} to X_i

$$\Rightarrow E(u_i^2) = \sigma^2 X_i$$
$$V(u_i)$$

$\Rightarrow$ transform as follows: (divide by $\sqrt{X_i}$)

$$\frac{Y_i}{\sqrt{X_i}} = B_1 \frac{1}{\sqrt{X_i}} + B_2 \sqrt{X_i} + V_i$$

$$Y_1^* = B_1^* + B_2 X_i^*$$

$$V_i = \frac{u_i}{\sqrt{X_i}}$$

Now $V(V_i) = \frac{1}{X_i} V(u_i) = \; = \frac{\sigma^2 X_i}{X_i} = \sigma^2$

$\Rightarrow$ can use OLS

Actually we are using WLS

In OLS we min $\sum e_i^2 = \sum (Y_i - b_1 - b_2 X_i)^2$

In WLS " " $\sum \left(\frac{e_i}{\sigma_i}\right)^2 = \sum \left[\frac{Y_i - b_1 - b_2 X_i}{\sigma_i}\right]^2$

Note: To est the transformed model, we must use regression thro the origin estimating procedure.

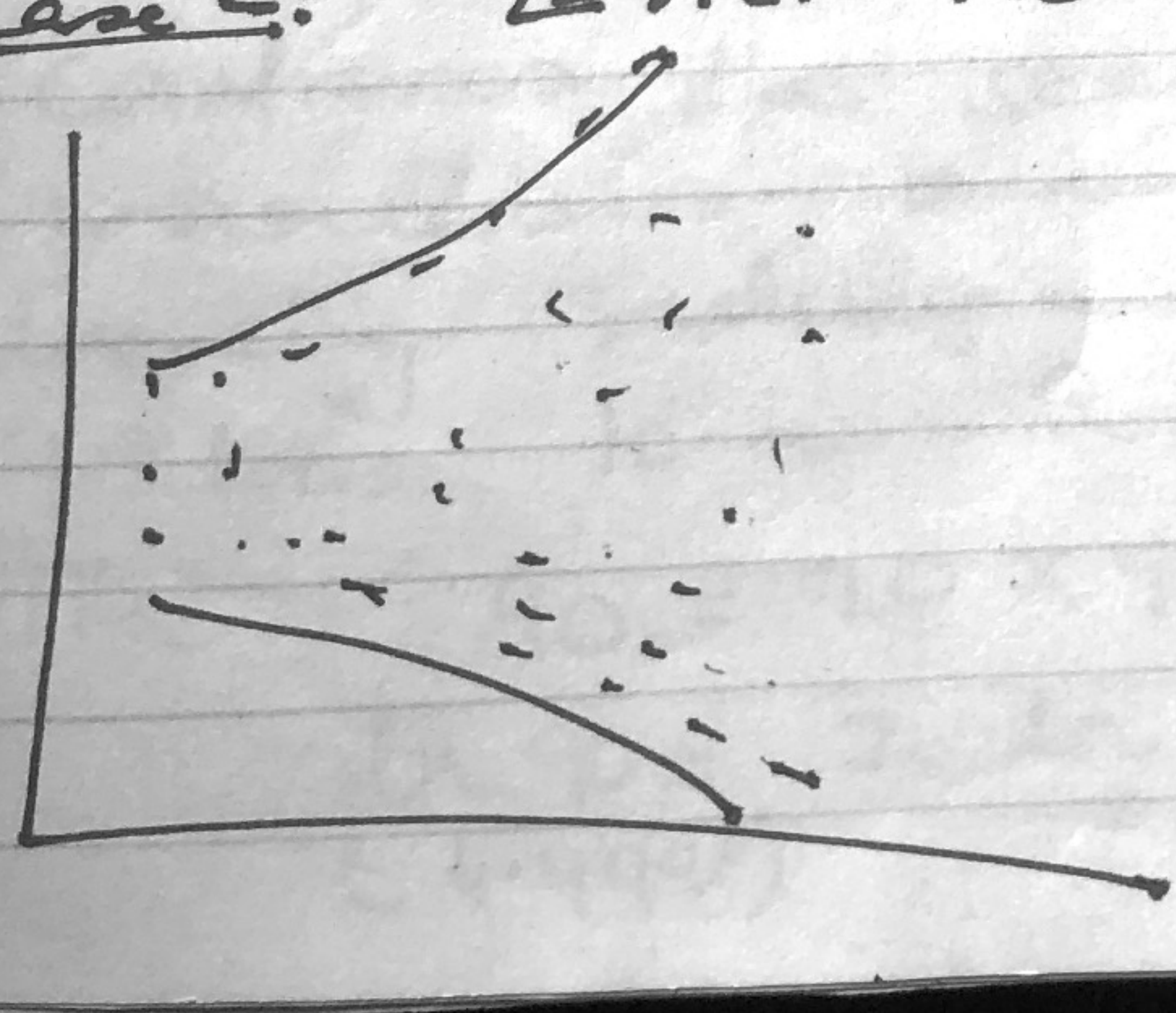
To get back to the original (untransformed) wage $\equiv^n$, just multiply both sides by $\sqrt{X_i}$

What if more than one EV in the model?

— May see which X_i is the most appropriate candidate for the transf^n

— or use $\hat{Y_i}$

Case 2: Error Variance prop^{al} to X_i^2



$\Rightarrow$ error variance increases proportionally to the square of X

$$E(u_i^2) = \sigma^2 X_i^2$$

So divide both sides by X_i.

$$\frac{Y_i}{X_i} = B_1 \frac{1}{X_i} + B_2 + \frac{u_i}{X_i}$$

$$V_i = \frac{u_i}{X_i}$$

$$V(v_i) = \frac{1}{X_i^2} V(u_i) = \frac{1}{\cancel{X_i^2}} \sigma^2 \cancel{X_i^2} = \sigma_i^2$$

Note slope coefficient becomes the intercept & the intercept becomes the slope coefficient

— but this Δ is only for estn, after estn multiply by X_i on both sides, we get the original model

<u>Respecification of the Model</u>

Estimating double log model instead of LIV model may eliminate or reduce HS.

→ This is so because the log transfn compresses the scales in which variables are measured thereby reducing a tenfold diff b/w 2 values to a 2 fold diff.

Thus $90 = 10 \times 9$

$\ln 90 = 2 \ln 9$
$(4.4998) \quad (2.1972)$

All the transformations discussed are called variance stabilizing transformations

— some amount of ad hoc use

— trial & error

— cd even transform the model by using a variable that was initially included but later removed from the model.

White's HS corrected std errors & t stats

In the presence of HS, OLS estr are inefficient.

⟹ conventionally computed std errors & t stats are suspect

White has developed an estn procedure that produces std errors of estimated regn coeff that take into a/c HS.

⟹ we can continue to use t & F tests except that they are now valid asymptotically

but these are valid only asymptotically

⟹ White's procedure does not Δ the values of the regn coeff but only their std errors.

Chow test - dummy variable interpretation.

In the saving income example used to explain the Chow test, we estimated three regressions — two of them constituting an unrestricted regression scenario & one being a restricted version.

So

$$\left.\begin{array}{l} Y_i = \lambda_1 + \lambda_2 X_i + u_i \\ Y_i = \gamma_1 + \gamma_2 X_i + u_i \end{array}\right\} \text{UR} \quad \begin{array}{l} 1970-81 \\ 1982-95 \end{array}$$

$$Y_i = \beta_1 + \beta_2 X_i + u_i \} \text{ RR } 1970-95$$

H_0: $\left.\begin{array}{l}\lambda_1 = \gamma_1 \\ \lambda_2 = \gamma_2\end{array}\right\}$ 2 restrictions

H_a: Not valid
$\Rightarrow$ either $\lambda_1 \neq \gamma_1$
or $\lambda_2 \neq \gamma_2$
or both.

$$F = \frac{RSS_{R} - RSS_{UR} / 2}{RSS_{UR} / n_1 + n_2 - k}$$

When H_0 gets rejected, it implies the restrictions are valid but we do not know whether the unrestricted regression make more sense because the

All of these are options that go with H_A

intercepts are diff (i.e $\lambda_1 \neq \gamma_1$ while $\lambda_2 = \gamma_2$) or the slopes alone are diff (i.e $\lambda_2 \neq \lambda_2$ while $\lambda_1 = \gamma_1$) or both are different (i.e $\lambda_1 \neq \gamma_1$ & $\lambda_2 \neq \lambda_2$)

To find if the intercepts are diff or the slopes are different we will have to do individual t tests.

or we need to introduce one restriction at a time. But that, it can be shown is equivalent to a t-test

2

This individual significance is better understood if we ~~t~~ consider the dummy variable interpretation of the Chow test

So consider the $=^n$

$$Y_t = \alpha_1 + \alpha_2 D_t + \beta_1 X_t + \beta_2 (D_t X_t) + u_t$$

Y – savings
X – income
t – time
D = 1 for 1982 - 1995
 = 0 for 1970 - 1981

Now

$$E(Y_t / D_t = 0) = \alpha_1 + \beta_1 X_t \qquad 1970\text{-}1981$$

$$E(Y_t / D_t = 1) = (\alpha_1 + \alpha_2) + (\beta_1 + \beta_2) X_t \qquad 1982\text{-}95$$

Notice that if we compare it to the earlier situation

$$\alpha_1 = \lambda_1 \quad \& \quad \alpha_1 + \alpha_2 = \gamma_1$$

$$\& \ \beta_1 = \lambda_2 \quad \& \quad \beta_1 + \beta_2 = \gamma_2$$

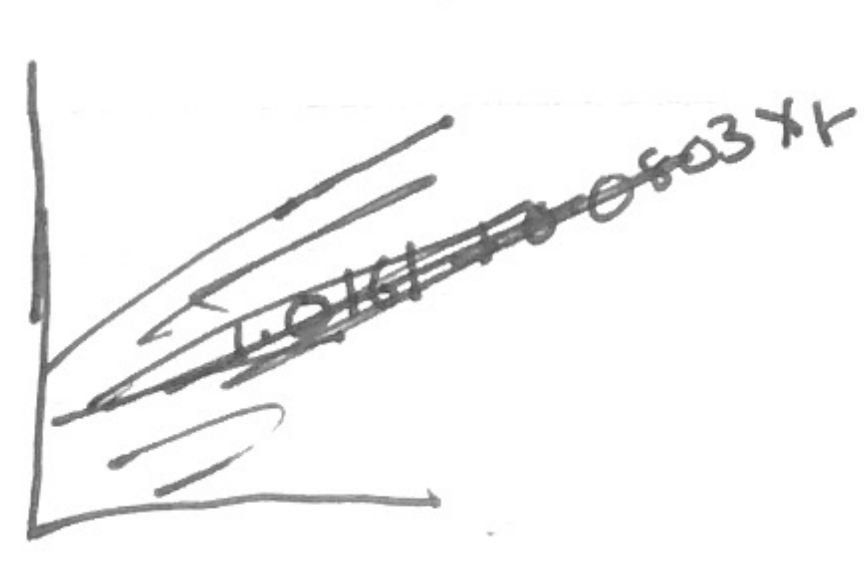
Now if this $=^n$ is est^ed to give the following results.

$$\hat{Y}_t = 1.0161 + 152.4786 D_t + 0.0803 X_t - 0.0655 D_t X_t$$

$$t \quad (0.0504) \quad (4.6090)^* \quad (5.5413)^* \quad (-4.0963)^*$$

D_t term: ↓ stat^y sig ⇓ diff in intercept

$D_t X_t$ term: ↓ stat^y sig ⇓ diff in slopes



If we wanted to use the restricted least squares version, then would not suffice to run one regression. First we would run the unrestricted reg^n i.e

$$Y_t = \alpha_1 + \alpha_2 D_t + \beta_1 X_t + \beta_2 (D_t X_t) + u_t$$

$\Rightarrow RSS_{UR}$

Impose one restriction

i.e $\alpha_2 = 0$

So restricted regression wd be

$$Y_t = \alpha_1 + \beta_1 X_t + \beta_2 D_t X_t + u_t$$

$\Rightarrow RSS_R$.

$$F = \frac{RSS_{UR} - RSS_R / 1}{RSS_{UR} / n-k} \sim F_{1, n-k}$$

Since $F_{1,n-k} = t^2_{n-k}$

this wd be equivalent to running just the unrestricted reg^n + carrying out a t test for $\alpha_2 = 0$

$H_0: \alpha_2 = 0 \qquad H_a: \alpha_2 \neq 0$

However when 2 restrictions are imposed

$\alpha_2 = 0 \quad + \quad \beta_2 = 0$

then will need to estimate the foll =^n as well

$Y_t = \alpha_1 + \beta_1 X_t + u_t$

$\Rightarrow RSS_R$

Now

$H_0: \alpha_2 = 0, \ \beta_2 = 0$ $\quad$ H_a: at least one is nonzero

$$F = \frac{RSS_{UR} - RSS_R / 2}{RSS_{UR} / n-k}$$

this is not equivalent to a t test

+ note that rejecting H_0 in favour of H_a will not tell you whether $\alpha_2 \neq 0$ or $\beta_2 \neq 0$ or both $\neq 0$.