Assessing the underlying (model-specific) assumptions

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

A B C D

A iid = independent & identically distributed

B N = normally distributed errors

C mean of distribution is 0
(guaranteed by the least squares estimation
— not really an assumption)

D constant variance σ^2 (homoscedasticity
or homoskedasticity)

We assess these assumptions using the residuals
(observed errors)

$$e_i = Y_i - \hat{Y}_i \qquad i = 1, 2, \ldots n$$

& we do this assessment using residual plots

Key assumptions (in order of importance)

1. errors are independent (no obvious pattern)
2. errors are identically distributed with constant variance σ^2
(homoscedastic errors)
3. errors are normally distributed

Use residual plots:

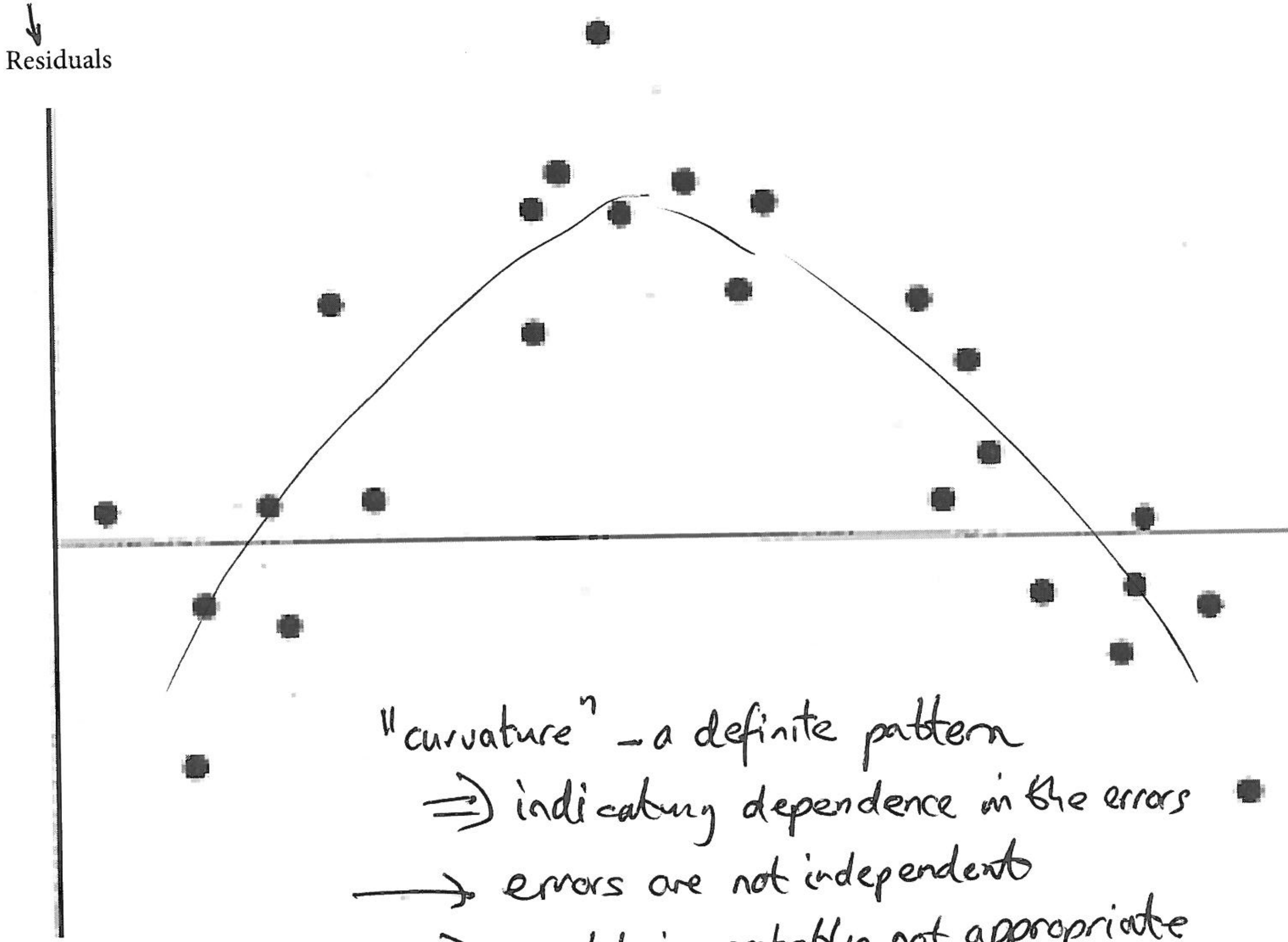
1 & 2 are best assessed using a plot of the (standardised) residuals vs fitted values
"main" residual plot

3 is best assessed using a normal quantile (qq plot) plot

other plots may be useful in diagnosing (getting more detail on) problems observed in the main residual plot (& occassionally the normal qq plot)

Plot I

$e_i = Y_i - \hat{Y}_i$



for SLR we could just use X_i

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$

MR

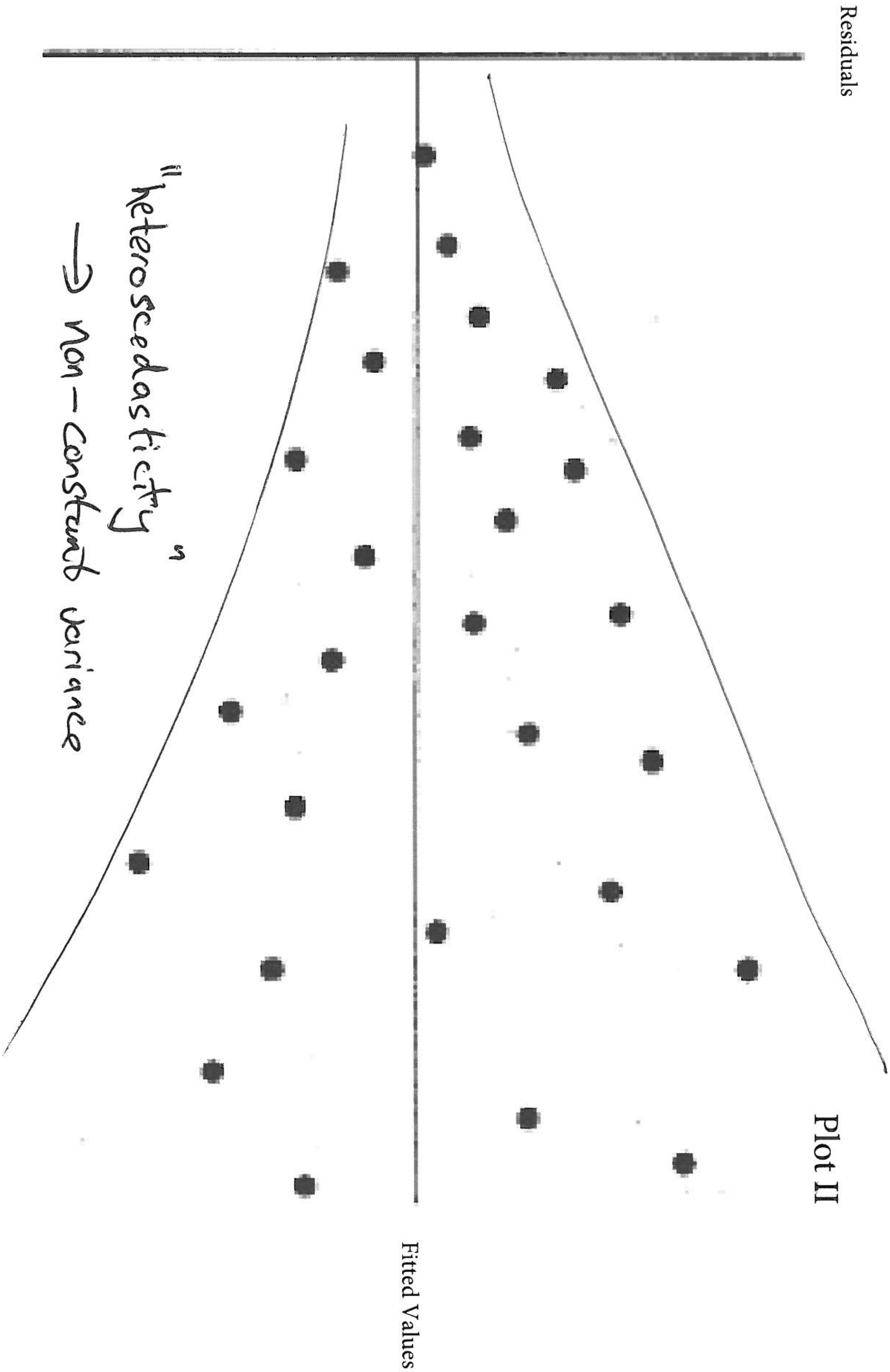
$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots \hat{\beta}_k X_{ki}$

"curvature" – a definite pattern

$\Rightarrow$ indicating dependence in the errors

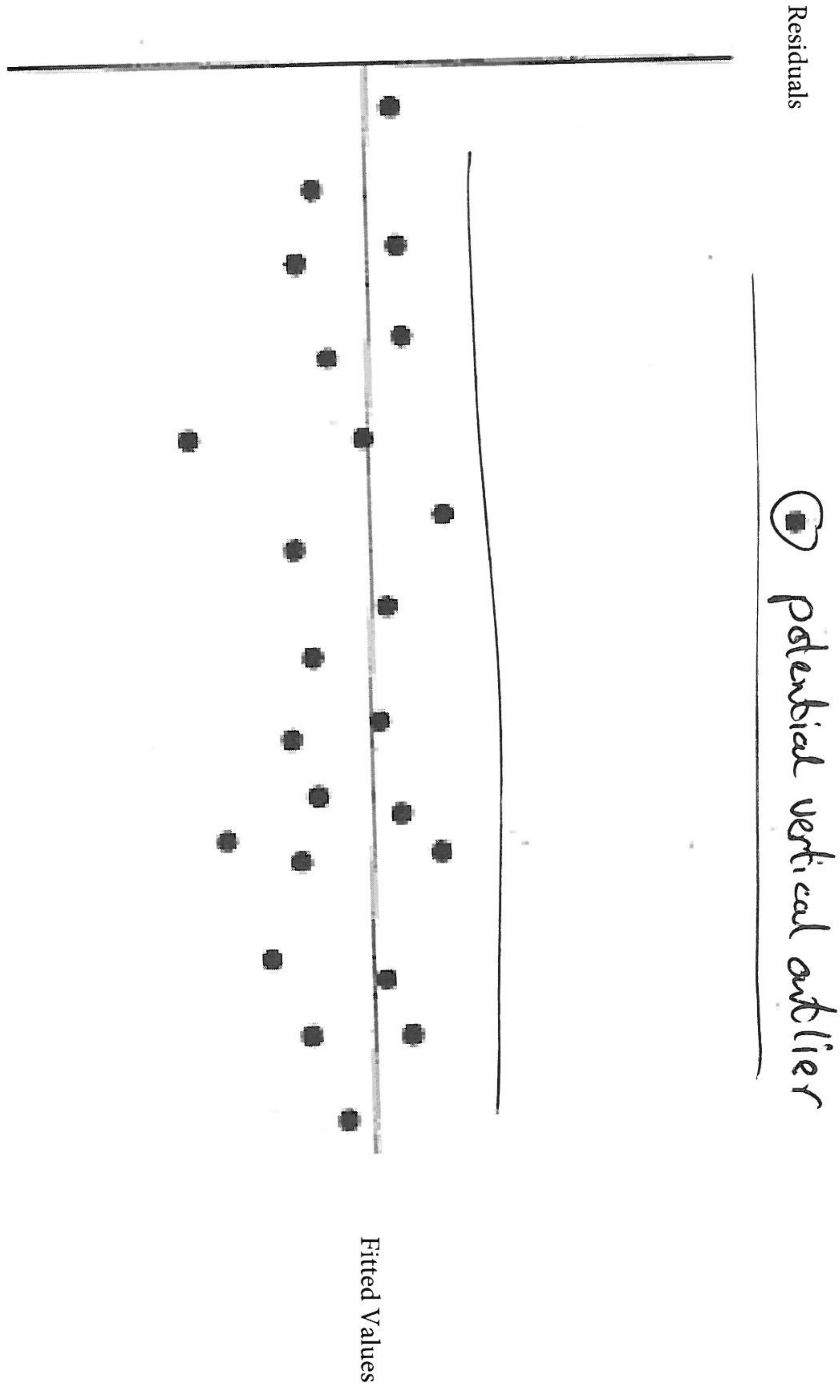
$\longrightarrow$ errors are not independent

$\longrightarrow$ model is probably not appropriate

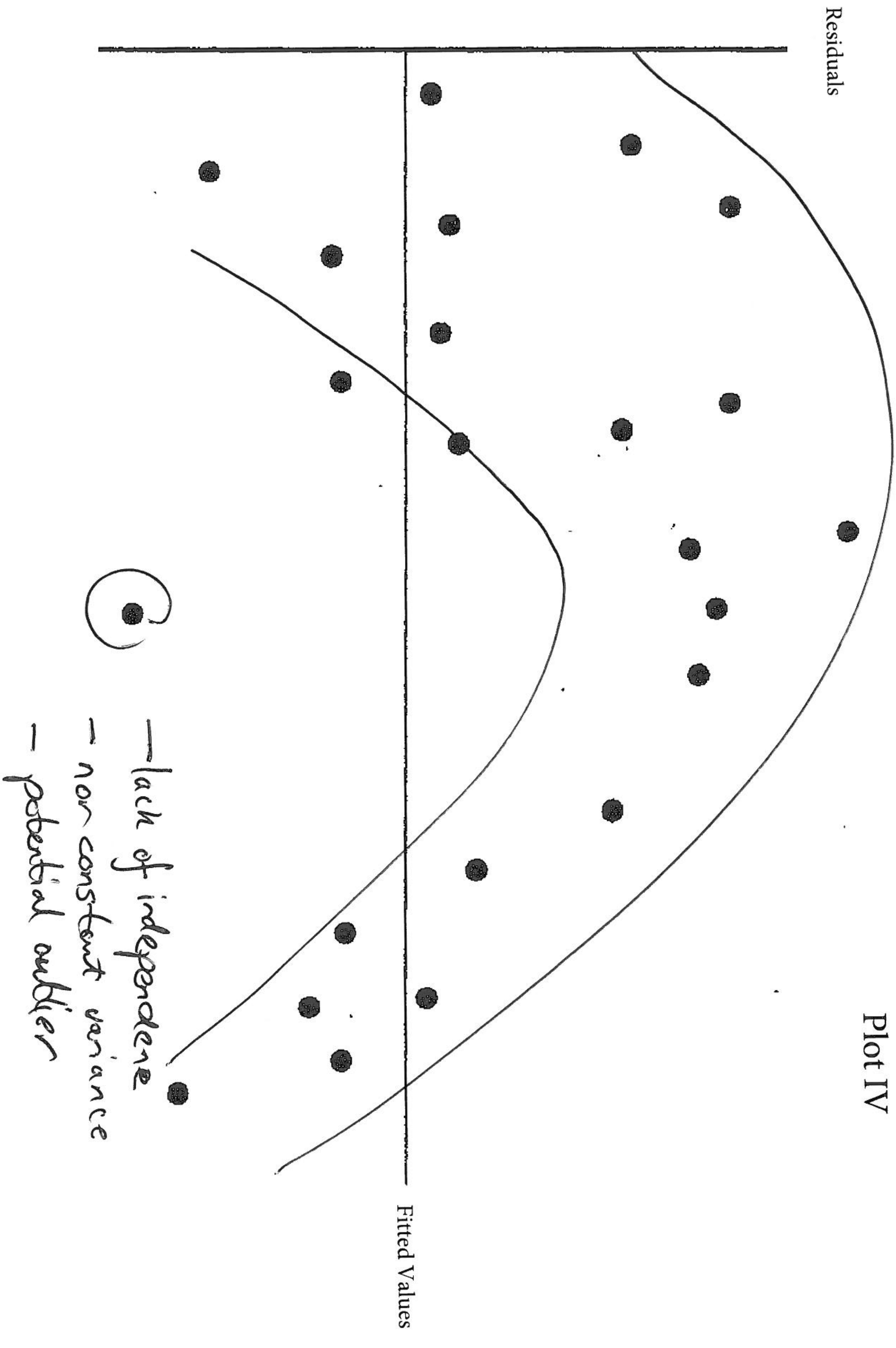


Plot II

"heteroscedasticity"
→ non-constant variance



Plot III


Residuals
Fitted Values
Plot IV
- lack of independence
- non constant variance
- potential outlier