WEIGHTED LEAST-SQUARES

1

CONSIDER the general linear model

$$Y_i = X_i\beta + \varepsilon_i \quad \text{~normal} \qquad (*)$$

where $E[\varepsilon_i] = 0$ for all i and $\text{Var}(\varepsilon_i) = \sigma_i^2$, $\left\{ i \in \{1, \dots, n\}, \; n = \text{\# of observations} \right.$

and the σ_i^2 can be different. Really, what we mean here is

$$E[\varepsilon_i \mid X_i] = 0 \text{ and } \text{Var}(\varepsilon_i \mid X_i) = \sigma_i^2,$$

so the variances of the errors can depend on values of the predictor variables. For now we assume the covariances are 0: $\left\{ i \in \{1, \dots, n\}, \; n = \text{\# of observations} \right.$

$$\text{Cov}(\varepsilon_i, \varepsilon_j \mid X_i, X_j) = 0 \text{ for } i \neq j.$$

Further, let W denote the inverse of the covariance matrix. Also define

$$W^{1/2} := \begin{bmatrix} \sqrt{w_1} & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{w_2} & 0 & \cdots & 0 \\ 0 & 0 & \sqrt{w_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sqrt{w_n} \end{bmatrix}$$

Note
$$\boxed{W^{-1} = \begin{bmatrix} 1/w_1 & 0 & 0 & \cdots & 0 \\ 0 & 1/w_2 & 0 & \cdots & 0 \\ 0 & 0 & 1/w_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}}$$

Here
$$W = \begin{bmatrix} w_1 & 0 & 0 & \cdots & 0 \\ 0 & w_2 & 0 & \cdots & 0 \\ 0 & 0 & w_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & w_n \end{bmatrix}$$
is KNOWN

Note $W^{1/2}$ is symmetric and $W^{1/2}W^{1/2} = W$. Also,

$$\cancel{(W^{1/2})^{-1}(W^{1/2})^{-1} = W} \quad (W^{-1})^{1/2}(W^{-1})^{1/2} = W^{-1},$$ and which also equals $(W^{\frac{1}{2}})^{-1}(W^{1/2})^{-1} = W^{-1}$ ... So we just write $W^{-1/2}W^{-1/2} = W^{-1}$

Now vectorize (*) and left-multiply by $W^{1/2}$:

$$\underbrace{W^{1/2}Y}_{=:\,Y^*} = \underbrace{W^{1/2}X}_{=:\,X^*}\beta + \underbrace{W^{1/2}\varepsilon}_{=:\,\varepsilon^*} \qquad (**)$$

So that $Y^* = X^*\beta + \varepsilon^*$. Now

$$E[\varepsilon^*] = E[W^{1/2}\varepsilon] = \underset{\uparrow\ \text{MATRIX of KNOWN CONSTANTS}}{W^{1/2}}E[\varepsilon] = 0$$

$$\underbrace{\sigma^2(\varepsilon^*)}_{\text{Covariance Matrix of the *'d errors}} = W^{1/2}\underbrace{\sigma^2(\varepsilon)}_{\text{Covariance matrix of original errors}}W^{1/2} = W^{1/2}W^{-1}W^{1/2}$$

$$= \underbrace{W^{1/2}W^{-1/2}}_{=I_{n\times n}}\underbrace{W^{-1/2}W^{1/2}}_{=I_{n\times n}}$$

Remember $\sigma^2(\varepsilon^*)$ and $\sigma^2(\varepsilon)$ are $n \times n$, n = # of observations

Thus, the model in (**) satisfies the mean = 0 and const. error variance assumptions for the general~~ized~~ linear model (also, normality of the ε^* is inherited from the normality of the ε). Thus

$$b^* = (X^{*T}X^*)^{-1}X^{*T}Y^*$$

contain the least-squares estimators for the β vector.

3

We can express this as

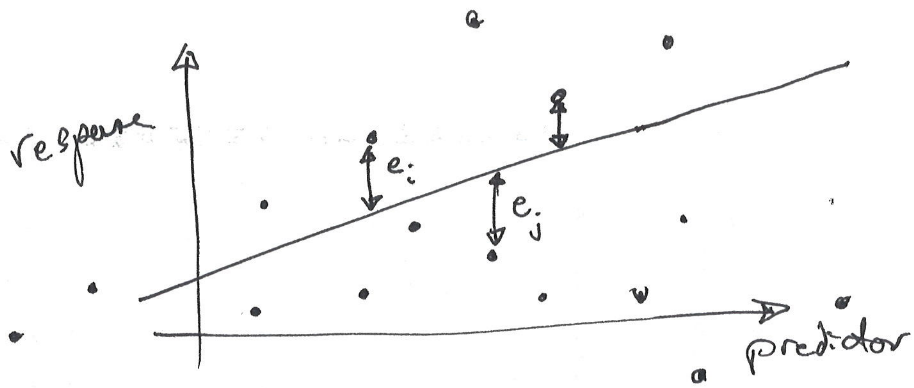
$$b^* = \Big[\underbrace{(W^{1/2}X)}_{=X^*}^T \underbrace{(W^{1/2}X)}\Big]^{-1}(W^{1/2}X)^T \underbrace{W^{1/2}Y}_{=Y^*}$$

$$= \left[X^T W X\right]^{-1} X^T W Y. \quad (***)$$

What if (and this is usually the case) the w-values are unknown? How should we estimate them?? Either estimate the standard deviation function with $\hat{s}_i$ or the variance function with $\hat{v}_i$, and set $\boxed{w_i = \frac{1}{\hat{s}_i^2}}$ or $\boxed{w_i = \frac{1}{v_i}}$ (these methods are NOT the same...)

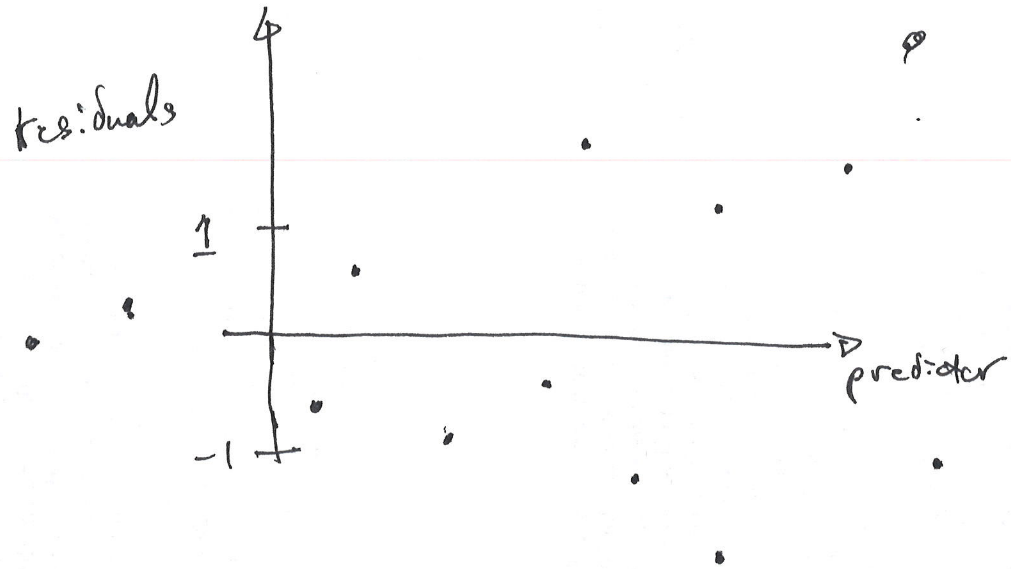
How to do this?

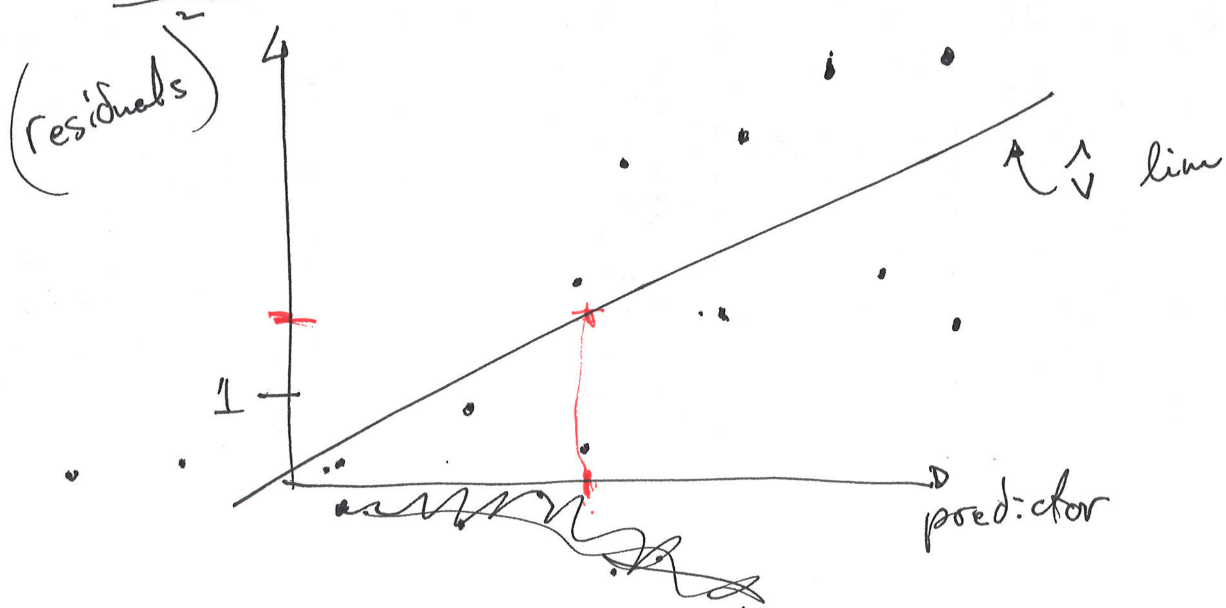
1. Fit regression model with unweighted least-squares (the usual way: $\hat{\beta} = (X^TX)^{-1}X^TY$).
2. Estimate the variance function w/ $\hat{v}$ or the std. dev. function $\hat{s}$ with the fits of a least squares regression of the squared errors or errors against the predictor variables of interest.
3. Use the fits from ② to get weights.
4. Use (***) and weights from ③ to get b^*.
5. Can iterate... start back @ ① with $\hat{\beta}_{new} = b^*$, get new residuals for ②, etc.



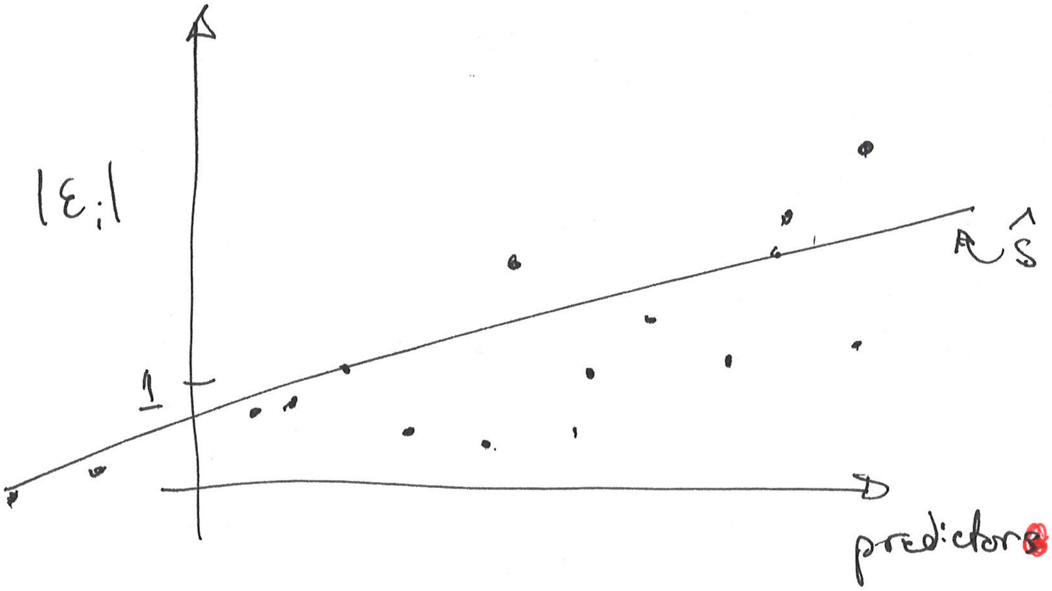
Recall $E[e_i] = 0$, and $Var(e_i) = E[e_i^2] - \underbrace{\left(E[e_i]\right)^2}_{=0} = E[e_i^2]$.

This is why we can think to model the variance of the errors from the squared residuals vs. predictors of interest.





Or, can model (w/ least-squares regression) the std. dev. of least-squares reg. line, regressing $|\varepsilon_i|$ vs. predictor(s):



Use $|\varepsilon_i|$ vs. preds or ε_i^2 vs. preds as appropriate.

Other than $\boxed{w_i = \frac{1}{\hat{s}_i^2}}$ or $\boxed{w_i = \frac{1}{v_i}}$ you can try other things: If the response (y_i is an average) of n_i observations, consider $w_i = n_i$. An alternative to these ... could try $w_i = \frac{1}{\hat{s}_i}$. Try something that comes from a good linear fit like this.

If two or more predictors (say x_1, x_2) contribute to the heteroskedasticity, try regressing ε_i^2 or $|\varepsilon_i|$ onto $x_1, x_2, x_1 x_2$, and maybe x_1^2 and x_2^2. Then build a regression model, get the fits, ...