Weighted Least Squares (WLS)

Intro. The **Weighted Least Squares (WLS)** method extends ordinary least squares by incorporating observation-specific weights. The basic model structure remains similar to OLS:

$$\hat{y}(\boldsymbol{x}) = \boldsymbol{x}^\intercal \hat{\boldsymbol{\beta}} + \varepsilon(\boldsymbol{x}) \tag{1}$$

where $\boldsymbol{x}$ is the vector of features, $\boldsymbol{\beta}$ is the vector of parameters, $y(\boldsymbol{x})$ is the target variable, and $\varepsilon(\boldsymbol{x})$ is the error term.

Heteroscedasticity. can be eliminated by applying weighted LS.

For a model with non-constant variance of the error term:

$$y = X\beta + \varepsilon, \quad \operatorname{Var}[\varepsilon(\boldsymbol{x})] = y(\boldsymbol{x})^2 \cdot \sigma^2$$

To apply WLS, the weights must have a negative square unit:

$$w(\boldsymbol{x}) = \frac{1}{y(\boldsymbol{x})^2}$$

This leads to the transformations:

$$\boldsymbol{y}' = \frac{\boldsymbol{y}}{\sqrt{w}}, \quad \boldsymbol{x}' = \frac{\boldsymbol{x}}{\sqrt{w}}, \quad \varepsilon' = \frac{\varepsilon}{\sqrt{w}}$$

The weight matrix is:

$$W = \begin{pmatrix} \frac{1}{y(\boldsymbol{x}_1)^2} & & \\ & \ddots & \\ & & \frac{1}{y(\boldsymbol{x}_\ell)^2} \end{pmatrix}$$

and

$$\boldsymbol{y}' = \sqrt{W}\boldsymbol{y}, \quad \boldsymbol{x}' = \sqrt{W}\boldsymbol{x}, \quad \varepsilon' = \sqrt{W}\varepsilon$$

Now, the model can be formulated as a homoscedastic least squares problem:

$$\boldsymbol{y}' = X'\beta + \varepsilon', \quad \operatorname{Var}[\varepsilon'(\boldsymbol{x})] = \sigma^2$$

- ✳ Each observation $\boldsymbol{x}$ has associated weights $w(\boldsymbol{x})$ that reflect the importance of that particular observation.
- ✳ This method minimizes the weighted sum of squared residuals:

$$\text{RSS} = \sum_{\boldsymbol{x} \in X^\ell} w(\boldsymbol{x}) \cdot (y(\boldsymbol{x}) - \hat{y}(\boldsymbol{x}))^2 \to \min_{\boldsymbol{\beta}}. \tag{2}$$

- ✳ The solution to this minimization problem is given by:

$$\boldsymbol{\beta}^* = \underbrace{(X^\intercal W X)^{-1} X^\intercal W}_{X_W^+}\boldsymbol{y}, \tag{3}$$

where W is the diagonal matrix of weights, and X_W^+ is the weighted pseudo-inverse.

Weight matrix. For a weighted

$$\text{RSS} = \sum_{\boldsymbol{x} \in X^\ell} w(\boldsymbol{x}) \cdot (y(\boldsymbol{x}) - \hat{y}(\boldsymbol{x}))^2 \tag{4}$$

let's introduce the weight matrix:

$$W := \operatorname{diag}(w(\boldsymbol{x}_1), \ldots, w(\boldsymbol{x}_\ell)) = \begin{pmatrix} w(\boldsymbol{x}_1) & & \\ & \ddots & \\ & & w(\boldsymbol{x}_\ell) \end{pmatrix} \tag{5}$$

Matrix form. Thus, we can rewrite the RSS in matrix form as a quadratic form:

$$\text{RSS} = (\boldsymbol{y} - X\boldsymbol{\beta})^\intercal W (\boldsymbol{y} - X\boldsymbol{\beta}). \tag{6}$$

Quadratic form. is a function of the form:

$$Q(x_1, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j} x_i x_j.$$

Coefficients $a_{i,j}$ can be arranged in a symmetric matrix A, and the quadratic form can be written in matrix form as:

$$Q(\boldsymbol{x}) = \boldsymbol{x}^T A \boldsymbol{x}.$$

Back to standard LS. The weighted LS problem can be easily reformulated as a standard LS problem by replacing the original variables with transformed ones:

$$\boldsymbol{y}' := W^{\frac{1}{2}}\boldsymbol{y}, \quad X' := W^{\frac{1}{2}}X, \quad \boldsymbol{\varepsilon}' := W^{\frac{1}{2}}\boldsymbol{\varepsilon} \tag{7}$$

Substituting these transformations into the original model, we get:

$$\boldsymbol{y}' = X'\boldsymbol{\beta} + \boldsymbol{\varepsilon}' \tag{8}$$

Analytical solution. Now, let's solve for $\boldsymbol{\beta}$ in the transformed model. Since W and $W^{\{\frac{1}{2}\}}$ are diagonal matrices, transposing them results in the same matrix:

$$\boldsymbol{\beta}^* = X'^{+}\boldsymbol{y}' = (X'^\intercal X')^{-1} X'^\intercal \boldsymbol{y}' \tag{9}$$

Expanding the expressions:

$$\begin{aligned} \boldsymbol{\beta}^* &= \left(\left(W^{\frac{1}{2}}X\right)^T W^{\frac{1}{2}}X\right)^{-1} \left(W^{\frac{1}{2}}X\right)^\intercal W^{\frac{1}{2}}\boldsymbol{y} \\ &= (X^\intercal W X)^{-1} X^\intercal W \boldsymbol{y} \end{aligned} \tag{10}$$

Therefore, the solution is:

$$\boldsymbol{\beta}^* = \underbrace{(X^\intercal W X)^{-1} X^\intercal W}_{X_W^+}\boldsymbol{y} \tag{11}$$