ELS Ex 7.5 page 259

For a linear smoother, that is $\underset{N\times 1}{\hat{y}} = \underset{N\times N \text{ symmetric matrix}}{S}\, \underset{N\times 1}{y}$

show that $\displaystyle\sum_{i=1}^{N} Cov(y_i, \hat{y}_i) = trace(S)\cdot \sigma_\varepsilon^2$

Where did σ_ε^2 come from? We assume an additive error model

$$Y_i = f(X_i) + \varepsilon_i \quad \text{where } E(\varepsilon_i) = 0,\ Var(\varepsilon_i) = \sigma_\varepsilon^2$$

and e_i and e_j independent

(Have we seen this before? Yes, in MLR the Hat matrix is the linear smoother: $\hat{\beta} = (X^TX)^{-1}X^Ty$ and $\hat{y} = X\hat{\beta} = \underbrace{X(X^TX)^{-1}X^T}_{H}y$

one possible version of S)

Start with:

$$\underset{N\times N \text{ matrix}}{Cov(\underset{N\times 1}{y}, \underset{N\times 1}{\hat{y}})} = \begin{bmatrix} Cov(y_1,\hat{y}_1) & Cov(y_1,\hat{y}_2) & \cdots & Cov(y_1,\hat{y}_N) \\ Cov(y_2,\hat{y}_1) & \ddots & & \\ \vdots & & \ddots & \\ Cov(y_N,\hat{y}_1) & & & Cov(y_N,\hat{y}_N) \end{bmatrix}$$

Then to get

$$\sum_{i=1}^{N} Cov(y_i,\hat{y}_i) = \underbrace{\text{sum of diagonal elements}}_{\text{def. of the trace}} \text{ of } Cov(y,\hat{y})$$

$$= trace(Cov(\underset{N\times 1}{y}, \underset{N\times 1}{\hat{y}}))$$

$$= \text{trace}\big(\text{Cov}(\underset{N\times 1}{y}, \overbrace{\underset{N\times N}{S}\,\underset{N\times 1}{y}}^{\hat{y}})\big) = \text{trace}\big(S \cdot \underbrace{\text{Cov}(y,y)}_{\text{Var}(y) \,=\, \sigma_\varepsilon^2 \cdot I}\big) = \text{trace}(S)\cdot \sigma_\varepsilon^2$$

[formulas from TMA4267]

since y_i, y_j independent

$$\text{trace}\left(\begin{bmatrix} S_{11} & \cdots & S_{1N} \\ \vdots & & \\ S_{N1} & \cdots & S_{NN}\end{bmatrix} \cdot \begin{bmatrix} \sigma_\varepsilon & 0 & \cdots & 0 \\ & \ddots & & \\ 0 & & & \sigma_\varepsilon \end{bmatrix}\right)$$

$$\text{Cov}(\underset{a\times 1}{V}, \underset{b\times 1}{W})_{a\times b} = E\big((V-\mu_V)(W-\mu_W)^T\big) \qquad [\text{here } a=b]$$

and if $W = SV$:

$$\mu_W = E(W) = E(SV) = SE(V) = S\mu_V$$

$$\text{Cov}(V, SV) = E\big((V-\mu_V)(SV-\mu_W)^T\big) = E\big((V-\mu_V)(SV-S\mu_V)^T\big)$$

$$= E\big((V-\mu_V)(V-\mu_V)^T S^T\big) = \text{Cov}(V)\cdot S^T$$

can be derived from Cov of linear combination

see for example

Härdle, Simar (2015) Applied Multivariate Statistical Analysis, Springer

(ebook at NTNU)

chapter 4.2 (link in L1.html) to book & notes

or classnotes from TMA4267

$$\text{trace}(AB^T) = \text{trace}(A^T B)$$

$$\text{trace}(\text{Cov}(V)\cdot S^T) = \text{trace}(S\,\text{Cov}(V))$$

symmetric