$$\underset{n\times 1}{y} = \underbrace{X}_{n\times(p+1)} \cdot \underset{(p+1)\times 1}{\beta} + \underset{n\times 1}{\varepsilon}$$

$$E[\varepsilon] = 0, \quad Var[\varepsilon] = \sigma^2 I_n$$

$$\hat\beta = (X^T X)^{-1} X^T y$$

$$E[\hat\beta] = E[(X^TX)^{-1}X^T \boxed{y}] \quad \text{(OLS estimator def)}$$

$$= E[(X^TX)^{-1}X^T (\boxed{X\beta + \varepsilon})] \quad \text{(linear model)}$$

$$= E[\underbrace{\boxed{(X^TX)^{-1}X^TX}}_{= I_{(p+1)}}\beta + (X^TX)^{-1}X^T\varepsilon]$$

$$= E[\beta + (X^TX)^{-1}X^T\varepsilon]$$

$$= \beta + E[(X^TX)^{-1}X^T\varepsilon] \quad (\beta \text{ is non-random})$$

$$= \beta + (X^TX)^{-1}X^T \underbrace{\boxed{E[\varepsilon]}}_{=0}$$

$$= \beta$$

summary:

$$\hat\beta = \beta + (X^TX)^{-1}X^T\varepsilon$$

$$Var[\hat\beta] = E\left[(\hat\beta - E[\hat\beta])(\hat\beta - E[\hat\beta])^T\right]$$

$$= E[(\hat\beta - \beta)(\hat\beta - \beta)^T] \quad \text{(unbiasedness of } \hat\beta\text{)} \quad E[\hat\beta] = \beta$$

$$= E\left[(X^TX)^{-1}X^T\varepsilon \underbrace{\boxed{[(X^TX)^{-1}X^T\varepsilon]^T}}_{\underset{1\times n}{\varepsilon^T} \underset{n\times(p+1)}{X} (X^TX)^{-1}}\right] \quad \text{(plug in partial result above)}$$

$$= E[(X^TX)^{-1}X^T\varepsilon\varepsilon^T X(X^TX)^{-1}]$$

$$= (X^TX)^{-1}X^T \underbrace{\boxed{E[\varepsilon\varepsilon^T]}}_{Var[\varepsilon] = \sigma^2_\varepsilon I_n} X(X^TX)^{-1} \quad \left(\text{scalar analogue is } E[axb] = aE[x]b\right)$$

$$= \sigma_\varepsilon^2 (X^T X)^{-1} \underbrace{\boxed{X^T X (X^T X)^{-1}}}_{I(p+1)}$$

Aside: $\mathrm{Var}[\varepsilon] = E[(\varepsilon - \underbrace{\boxed{E[\varepsilon]}}_{=0})(\varepsilon - \underbrace{\boxed{E[\varepsilon]}}_{=0})^T]$

$$= E[\varepsilon \varepsilon^T]$$