Ex 2.5 (a)

$$EPE(x_0) = E_{y_0|x_0} E_\tau (y_0 - \hat{y}_0)^2$$
$$= E_{y_0|x_0} E_\tau (y_0 - x_0^T\beta + x_0^T\beta - \hat{y}_0)^2 \quad \text{(trick)}$$

$$= \overset{①}{E_{y_0|x_0} E_\tau (y_0 - x_0^T\beta)^2} + \overset{②}{E_{y_0|x_0} E_\tau (x_0^T\beta - \hat{y}_0)^2}$$
$$+ \overset{③}{E_{y_0|x_0} E_\tau [2(y_0 - x_0^T\beta)(x_0^T\beta - \hat{y}_0)]}$$

① Note for any point i we have $y_i = x_i^T\beta + \varepsilon_i$

$\Rightarrow y_0 - x_0^T\beta = \varepsilon_0 \quad (*)$

$$E_{y_0|x_0} E_\tau (\varepsilon_0)^2 = \sigma^2 \xrightarrow{\text{using}} \begin{aligned} &Var(\varepsilon) = E(\varepsilon^2) - E(\varepsilon)^2 \\ &\Rightarrow E(\varepsilon^2) = Var(\varepsilon) + E(\varepsilon)^2 \\ &\Rightarrow E(\varepsilon^2) = \sigma^2 + 0 \end{aligned}$$

③ again using $(*)$ we get:

$$2 E_{y_0|x_0} E_\tau [(\varepsilon_0)(x_0^T\beta - \hat{y}_0)]$$

as $\varepsilon \sim$ iid

$$= 2 E_{y_0|x_0} E_\tau (\varepsilon_0) \, E_{y_0|x_0} E_\tau (x_0^T\beta - \hat{y}_0) = 0$$

② $E_{y_0|x_0} E_\tau (x_0^T\beta - \hat{y}_0)^2 = E_\tau (x_0^T\beta - \hat{y}_0)^2$ (No dependance on y_0)

$$= E_\tau (x_0^T\beta - E(\hat{y}_0) + E(\hat{y}_0) - \hat{y}_0)^2 \quad \text{(trick)}$$

again split into 3 expressions:

$$= \underset{ⓐ}{E_\tau (x_0^T\beta - E_\tau(\hat{y}_0))^2} + \underset{ⓑ}{E_\tau (E_\tau(\hat{y}_0) - \hat{y}_0)^2} + \overset{ⓒ}{2E_\tau[(x_0^T\beta - E_\tau(\hat{y}_0))(E_\tau(\hat{y}_0) - \hat{y}_0)]}$$

ⓐ $E_\tau (x_0^T\beta - E_\tau(\hat{y}_0))^2$

$= (x_0^T\beta - E_\tau(\hat{y}_0))^2$ (constant)

Note: $x_0^T\beta = E_\tau(\hat{y}_0) \quad (**)$

$= 0$

ⓒ again using $(**)$ we find:

$$2 E_\tau [(x_0^T\beta - E_\tau(\hat{y}_0))(E_\tau(\hat{y}_0) - \hat{y}_0)]$$

$= 0$

⑥ $E_\tau(E_\tau(\hat{y}_0) - \hat{y}_0)^2$

$= Var_\tau(\hat{y}_0)$ (by definition of Variance)

$= Var_\tau(x_0^T \hat{\beta})$

$= x_0^T Var_\tau(\hat{\beta}) x_0$

$= x_0^T (X^T X)^{-1} \sigma^2 x_0$ (using 3.8)

and in order to drop the assumption that the x_i are fixed (non-random) we simply replace with the expectation.

$= x_0^T E_\tau[(X^T X)^{-1}] x_0 \sigma^2$

Now putting all the pieces together we obtain that:

$$EPE(x_0) = E_{y_0|x_0} E_\tau (y_0 - \hat{y}_0)^2$$
$$= \sigma^2 + x_0^T E_\tau[(X^T X)^{-1}] x_0 \sigma^2$$

Ex 2.5 (b)

Given $EPE(x_0) = \sigma^2 + E_\tau x_0^T (X^T X)^{-1} x_0 \sigma^2$

We wish to find $E_{x_0} EPE(x_0)$

Assuming $E(X) = 0$ then $X^T X \to N\,cov(X)$

$$\Rightarrow E_{x_0} EPE(x_0) = \sigma^2 + \frac{E_{x_0} x_0^T\, cov(X)^{-1} x_0 \sigma^2}{N}$$

Now, we are told that we wish to get this equation in terms of the trace. Thus we will use the identity:

$$E(B^T A B) = trace(A\,cov(B)) + E(B)^T A E(B)$$

Where $A = P \times P$ and $B = P \times 1$ dimensions

Now,

$$E_{x_0} EPE(x_0) = \sigma^2 + [trace(cov(X)^{-1} cov(x_0)) + E(x_0)^T cov(X)^{-1} E(x_0)]\sigma^2/N$$

as $E(x_0) = 0$

$$= \sigma^2 + trace(cov(X)^{-1} cov(x_0))\sigma^2/N$$

$$= \sigma^2 + trace(I_p)\sigma^2/N$$

$$= \sigma^2 + \frac{p\sigma^2}{N}$$