1.

A. Now $\bar{\sigma}_F^2 \equiv (n-p)^{-1}\sum_{i=1}^n \hat{\epsilon}_i^2$ is an unbiased estimator of σ_F^2 if $rank(X) \equiv rank(C) = p$ --- here and below I'll use $\mathbf{X} \equiv X$ instead of C for the design matrix. Hence the "bias - corrected" residuals $\tilde{\epsilon}_i$ satisfy

$$E(\frac{1}{n}\sum_{i=1}^n \tilde{\epsilon}_i^2) = E(\frac{1}{n}\sum_{i=1}^n \epsilon^2) = \sigma_F^2 \, .$$

B. Now with $\underline{Y}^* = \mathbf{X}\underline{\hat{\beta}} + \underline{\hat{\epsilon}}^*$ and $\underline{\hat{\beta}}^* = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\underline{Y}^*$,

$$\begin{aligned} Cov_*(\underline{\hat{\beta}}^*) &= E_*(\underline{\hat{\beta}}^* - E_*\underline{\hat{\beta}}^*)(\underline{\hat{\beta}}^* - E_*\underline{\hat{\beta}}^*)^T \\ &= (X^TX)^{-1}X^T Cov_*(\underline{Y}^*)X(X^TX)^{-1} \\ &= (X^TX)^{-1}X^T Cov_*(\underline{\hat{\epsilon}}^*)X(X^TX)^{-1} \\ &= (X^TX)^{-1}\,\hat{\sigma}_F^2 \end{aligned}$$

with $\hat{\sigma}_F^2 \equiv n^{-1}|\underline{\hat{\epsilon}} - n^{-1}\underline{1}^T\underline{\hat{\epsilon}}|^2$, since

$$Cov_*(\underline{\hat{\epsilon}}^*) = I\,\frac{1}{n}\sum_{i=1}^n(\hat{\epsilon}_i - \frac{1}{n}\sum_{j=1}^n \hat{\epsilon}_j)^2 = I\,\frac{1}{n}\,|\underline{\hat{\epsilon}} - n^{-1}\underline{1}^T\underline{\hat{\epsilon}}|^2 \equiv I\,\hat{\sigma}_F^2 \, .$$

If the design matrix X contains a column of 1's (so the regression model contains an intercept or constant term), then $\underline{\hat{\epsilon}} \perp \underline{1}$ and this reduces to the expression for $\hat{\sigma}_F^2$ given in E&T, page 109, (9.18). It seems to me that E&T have implicitly assumed this to be the case in deriving their (9.18) -- since I do not see the assumption explicitly in their development. (This is related to the point of the next problem, problem 5.2, in this problem set -- an issue which they seem to have ignored.)

If we sample instead from the "bias - corrected" residuals $\tilde{\epsilon}_i$, then with $\underline{Y}^* = \mathbf{X}\underline{\hat{\beta}} + \underline{\tilde{\epsilon}}^*$ and $\underline{\hat{\beta}}^*$ defined in terms of $\underline{Y}^*$ as before,

$$\begin{aligned} Cov_*(\underline{\hat{\beta}}^*) &= E_*(\underline{\hat{\beta}}^* - E_*\underline{\hat{\beta}}^*)(\underline{\hat{\beta}}^* - E_*\underline{\hat{\beta}}^*)^T \\ &= (X^TX)^{-1}X^T Cov_*(\underline{Y}^*)X(X^TX)^{-1} \\ &= (X^TX)^{-1}X^T Cov_*(\underline{\tilde{\epsilon}}^*)X(X^TX)^{-1} \\ &= (X^TX)^{-1}\,\bar{\sigma}_F^2 \end{aligned}$$

with $\bar{\sigma}_F^2 \equiv (n-p)^{-1}|\underline{\hat{\epsilon}} - n^{-1}\underline{1}^T\underline{\hat{\epsilon}}|^2$ since

$$Cov_*(\underline{\hat{\epsilon}}^*) = I\,\frac{1}{n}\sum_{i=1}^{n}(\tilde{\epsilon}_i - \frac{1}{n}\sum_{j=1}^{n}\tilde{\epsilon}_j)^2 = I\,\frac{1}{n-p}\,|\underline{\hat{\epsilon}} - n^{-1}\underline{1}^T\underline{\hat{\epsilon}}|^2 \equiv I\bar{\sigma}_F^2 .$$

2. A. Now $\hat{Y} = X\hat{\beta} = X(X^TX)^{-1}X^TY = HY$ so $\underline{\hat{\epsilon}} = Y - \hat{Y} = (I-H)Y$ and

$$\underline{\hat{\epsilon}} - \underline{\epsilon} = (I-H)Y - (Y - X\underline{\beta}) \tag{a}$$

$$= -H(X\underline{\beta} + \underline{\epsilon}) + X\underline{\beta} = -H\underline{\epsilon}$$

since $H(X\underline{\beta}) = X\underline{\beta}$ ($X\underline{\beta}$ is already in the column space of X !)

B. Let $\hat{\epsilon}_i^*$ be a sample with replacement from $\{\hat{\epsilon}_i : i = 1, \cdots, n\}$, and let $Y_i^* = x_i\hat{\beta} + \hat{\epsilon}_i$, $i = 1, \cdots, n$. Thus in vector notation, $Y^* = X\underline{\hat{\beta}} + \underline{\hat{\epsilon}}^*$ and

$$\underline{\hat{\beta}}^* = (X^TX)^{-1}X^TY^* = (X^TX)^{-1}X^T(X\underline{\hat{\beta}} + \underline{\hat{\epsilon}}^*) \tag{b}$$

$$= \underline{\hat{\beta}} + (X^TX)^{-1}X^T\underline{\hat{\epsilon}}^* .$$

Thus

$$\sqrt{n}(\underline{\hat{\beta}}^* - \underline{\hat{\beta}}) = (\frac{1}{n}X^TX)^{-1}(\frac{1}{n}X^T)\sqrt{n}\underline{\hat{\epsilon}}^* , \tag{c}$$

and, since $E_*(\hat{\epsilon}_i^*) = \frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i$, the expected value is

$$E_*(\sqrt{n}(\underline{\hat{\beta}}^* - \underline{\hat{\beta}})) = (\frac{1}{n}X^TX)^{-1}(\frac{1}{n}X^T)\underline{1}\,\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\epsilon}_i \tag{d}$$

$$= (\frac{1}{n}X^TX)^{-1}(\frac{1}{n}X^T)\underline{1}\,Z_n .$$

C. To show that

$$\sqrt{n}(\underline{\hat{\beta}} - \underline{\beta}) \to_d N_p(0, \sigma^2V^{-1}) , \tag{e}$$

first write $\sqrt{n}(\underline{\hat{\beta}} - \underline{\beta}) = \sqrt{n}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\underline{\epsilon}$ so that, for any fixed vector $\lambda \in R^p$,

$$\underline{\lambda}^T\sqrt{n}(\hat{\beta} - \beta) = \sqrt{n}\underline{\lambda}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\underline{\epsilon} \equiv \sum_{i=1}^{n}a_{ni}\,\epsilon_i \equiv \sum_{i=1}^{n}X_{ni}$$

where the vector $\underline{a}_n \equiv \sqrt{n}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\underline{\lambda}$. Hence we have $EX_{ni} = 0$, $Var(X_{ni}) = a_{ni}^2\sigma^2$, and

$$\sigma_n^2 \equiv \sum_{i=1}^{n} \sigma_{ni}^2 = \sigma^2 \, |\underline{a}|^2$$

$$= \sigma^2 \, n \underline{\lambda}^T (\mathbf{X}^T \mathbf{X})^{-1} \underline{\lambda}$$

$$\rightarrow \sigma^2 \underline{\lambda}^T V^{-1} \underline{\lambda} > 0 \qquad \text{(f)}$$

since V is positive definite.

To check the Lindeberg condition, write

$$\frac{1}{\sigma_n^2} \sum_{i=1}^{n} E|X_{ni}|^2 1_{[|X_{ni}| > \epsilon\sigma_n]}$$

$$= \frac{1}{\sigma_n^2} \sum_{i=1}^{n} a_{ni}^2 \, E\{\epsilon_1^2 1_{[|\epsilon_1| > \epsilon\sigma_n/|a_{ni}|]}\}$$

$$\leq \frac{1}{\sigma^2} E\{\epsilon_1^2 1_{[|\epsilon_1| > \epsilon\sigma_n/\max_i |a_{ni}|]}\}$$

$$\rightarrow 0 \qquad \text{by the DCT since} \quad E(\epsilon^2) < \infty$$

if $\max_{1 \leq i \leq n} |a_{ni}| \rightarrow 0$. But we can write $a_{ni} = \sqrt{n} \underline{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \underline{\lambda}$, so that, by Cauchy - Schwarz,

$$\max_{1 \leq i \leq n} |a_{ni}^2| \leq n \max_{1 \leq i \leq n} (\underline{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \underline{x}_i) (\underline{\lambda}^T (\mathbf{X}^T \mathbf{X})^{-1} \underline{\lambda})$$

$$= \max_{1 \leq i \leq n} h_{ii} \, \underline{\lambda}^T (\frac{1}{n} \mathbf{X}^T \mathbf{X})^{-1} \underline{\lambda}) \rightarrow 0 \, \underline{\lambda}^T V^{-1} \underline{\lambda} = 0 \, .$$

Hence, by the Lindeberg - Feller CLT and (f)

$$\underline{\lambda}^T \sqrt{n} (\underline{\hat{\beta}} - \underline{\beta}) \rightarrow_d N_p(0 \, , \underline{\lambda}^T V^{-1} \underline{\lambda} \sigma^2) \, .$$

By the Cramér - Wold device, this yields (e) under the hypothesis $\max_i h_{ii} \rightarrow 0$.

D. To calculate the variance, we first note that from A, $\hat{\epsilon} = (I - H)\epsilon$, so $E(\hat{\epsilon}) = (I - H) E(\epsilon) = 0$ and $E(Z_n) = 0$. Similarly,

$$Var(Z_n) = \frac{1}{n} (\underline{1}^T (I - H)(I - H) \underline{1}) \, \sigma^2 \qquad \text{(g)}$$

$$= \sigma^2 \{1 - (\frac{1}{n} \underline{1}^T X)(\frac{1}{n} X^T X)^{-1} (\frac{1}{n} X^T \underline{1})\} \, .$$

E. If $\frac{1}{n} X^T \underline{1} \rightarrow h$, $\frac{1}{n} X^T X \rightarrow V$ with V positive definite, and

$h^T V^{-1} h \ < \ 1$, then

$$Var(Z_n) \ = \ \sigma^2 \{1 \ - \ (\frac{1}{n} \underline{1}^T X)(\frac{1}{n} X^T X)^{-1} (\frac{1}{n} X^T \underline{1})\} \ .$$

$$\rightarrow \ \sigma^2 \{1 \ - \ h^T V^{-1} h\} \ \equiv \ \sigma^2 c^2 \ > \ 0 \ .$$

To show that $Z_n \ \rightarrow_d$ using the Lindeberg - Feller CLT requires a bit more in the way of hypotheses and some more work: Write

$$Z_n \ = \ n^{-1/2} \underline{1}^T (I \ - \ H) \underline{\epsilon} \ \equiv \ \sum_{i=1}^{n} c_{ni} \, \epsilon_i \ \equiv \ \sum_{i=1}^{n} X_{ni} \ .$$

so that the vector $\underline{c}_n \ = \ n^{-1/2} (I - H) \underline{1}$. Thus $E \, X_{ni} \ = \ 0$, $\sigma_{ni}^2 \ = \ Var(\, X_{ni}) \ = \ c_{ni}^2 \sigma^2$, and, as above,

$$\sigma_n^2 \ = \ \sum_{i=1}^{n} \sigma_{ni}^2 \ = \ \sigma^2 \sum_{i=1}^{n} c_{ni}^2 \ = \ \sigma^2 \, \underline{c}^T \underline{c}$$

$$= \ \sigma^2 \, (1 \ - \ \frac{1}{n} \underline{1}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \underline{1}) \ \ \rightarrow \ \sigma^2 (1 \ - \ \underline{h}^T V^{-1} \underline{h})$$

under the above hypotheses. Finally, if $\max_{1 \, \leq \, i \, \leq \, n} |c_{ni}| \ \rightarrow \ 0$, then, for $\epsilon \ > \ 0$

$$\frac{1}{\sigma_n^2} \sum_{i=1}^{n} E |X_{ni}|^2 1_{[|X_{ni}| \, > \, \epsilon \sigma_n]}$$

$$= \ \frac{1}{\sigma_n^2} \sum_{i=1}^{n} c_{ni}^2 \, E \{ \epsilon_1^2 1_{[|\epsilon_1| \, > \, \epsilon \sigma_n / |c_{ni}|]} \}$$

$$\leq \ \frac{1}{\sigma^2} \, E \{ \epsilon_1^2 1_{[|\epsilon_1| \, > \, \epsilon \sigma_n / \max_i |c_{ni}|]} \}$$

$$\rightarrow \ 0 \qquad \text{by the DCT since} \quad E(\epsilon^2) \ < \ \infty \tag{h}$$

if $\max_{1 \, \leq \, i \, \leq \, n} |c_{ni}| \ \rightarrow \ 0$. But

$$|c_{ni}| \ \leq \ n^{-1/2} \ + \ n^{-1/2} | \sum_{j=1}^{n} h_{ij} |$$

$$\leq \ n^{-1/2} \ + \ \sqrt{n^{-1} \sum_{j=1}^{n} h_{ij}^2}$$

$$= \ n^{-1/2} \ + \ \sqrt{n^{-1} h_{ii}}$$

$$\leq n^{-1/2} + n^{-1/2} \leq 2n^{-1/2} \to 0$$

since $H = H^T$, $HH = H$, and the diagonal elements of H satisfy $0 \leq h_{ii} \leq 1$. Hence the Lindeberg - Feller CLT yields

$$\frac{Z_n}{\sigma_n} \to N(0,1) ;$$

combining this with (d) yields

$$\sqrt{n} E_*(\hat{\beta}^* - \hat{\beta}) \to_d V^{-1} h\, N(0, c^2\sigma^2) . \tag{i}$$

We conclude from (e) and (i) that the bootstrap *fails* in this situation (at least under the additional hypothesis that $\max_i h_{ii} \to 0$).

3. When the resampling is done from the *centered* residuals $\underline{\hat{\epsilon}} - \underline{1}(\underline{1}^T \underline{\hat{\epsilon}}/n)$, the nonzero term in $E_*(\sqrt{n}(\underline{\hat{\beta}}^* - \underline{\hat{\beta}}))$ which we investigated in problem 2 above vanishes: Since

$$\sqrt{n}(\underline{\hat{\beta}}^* - \underline{\hat{\beta}}) = (X^T X)^{-1} X^T \underline{\hat{\epsilon}}^* ,$$

where

$$E_*(\underline{\hat{\epsilon}}^*) = \underline{1} \frac{1}{n} \sum_{i=1}^{n} (\hat{\epsilon}_i - \underline{1}^T \underline{\hat{\epsilon}}/n) = \underline{0} ,$$

it follows that

$$E_*\{\sqrt{n}(\hat{\beta}^* - \hat{\beta})\} = (X^T X)^{-1} X^T E_* \underline{\hat{\epsilon}}^* = \underline{0} .$$

Furthermore,

$$\begin{aligned} E_*\{[\sqrt{n}(\underline{\hat{\beta}}^* - \underline{\hat{\beta}})]^{\otimes 2}\} &= n (X^T X)^{-1} X^T\, E_*(\underline{\hat{\epsilon}}^* \underline{\hat{\epsilon}}^{*T}) X (X^T X)^{-1} \\ &= n (X^T X)^{-1} \hat{\sigma}_F^2 \end{aligned}$$

since

$$E_*(\underline{\hat{\epsilon}}^* \underline{\hat{\epsilon}}^{*T}) = I \frac{1}{n} \sum_{i=1}^{n} (\hat{\epsilon}_i - \underline{1}\underline{\hat{\epsilon}}/n)^2 \equiv \hat{\sigma}_F^2 I .$$

This modification of the bootstrap procedure seems appropriate when the design matrix X does not contain a column of 1 's. See Freedman (1981), *Ann. Statist.* **9,** 1218 - 1228; especially the discussion on page 1220, the (positive!) theorem on page 1223, and the discussion on page 1224 (upon which this problem is based). It is a bit surprising that Efron and Tibshirani make no mention of this modification in their chapter 9.