Solution (Exercise 7.1)

Here, I will derive 7.1 (iii)-(vi)
We have already shown (in class) that we can write $\widehat{\varepsilon}_i$ as follows:

$$\begin{aligned}\widehat{\varepsilon}_i &= (\varepsilon_i - \overline{\varepsilon}) - \left(X_i - \overline{X}\right) A \\ \text{where } A &= \frac{\sum_{k=1}^n \varepsilon_k \left(X_k - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}.\end{aligned}$$

We have also shown that

$$\begin{aligned}E(A) &= 0, \\ Var(A) &= \frac{\sigma^2}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}.\end{aligned}$$

.

(iii)

(0.1)
$$\begin{aligned}Var(\widehat{\varepsilon}_i) &= Var\left((\varepsilon_i - \overline{\varepsilon}) - \left(X_i - \overline{X}\right) A\right) \\ &= Var(\varepsilon_i - \overline{\varepsilon}) + Var\left(\left(X_i - \overline{X}\right) A\right) - 2\left(X_i - \overline{X}\right) Cov(A, \varepsilon_i - \overline{\varepsilon})\end{aligned}$$

$$\begin{aligned}Var(\varepsilon_i - \overline{\varepsilon}) &= Var\left(\varepsilon_i - \frac{\sum_k \varepsilon_k}{n}\right) \\ &= Var(\varepsilon_i) + Var\left(\frac{\sum_k \varepsilon_k}{n}\right) - 2Cov\left(\varepsilon_i, \frac{\sum_k \varepsilon_k}{n}\right) \\ &= \sigma^2 + \frac{n\sigma^2}{n^2} - 2Cov\left(\varepsilon_i, \frac{\varepsilon_i}{n}\right), \text{ note } Cov(\varepsilon_i, \varepsilon_j) = 0 \text{ for } i \neq j \\ &= \sigma^2 + \frac{\sigma^2}{n} - \frac{2\sigma^2}{n} = \sigma^2\left(1 - \frac{1}{n}\right)\end{aligned}$$

$$\begin{aligned}Var\left(\left(X_i - \overline{X}\right) A\right) &= \left(X_i - \overline{X}\right)^2 Var(A) \\ &\quad \text{(Recall our derivation of } var(A) \text{ that we did in class)} \\ &= \left(X_i - \overline{X}\right)^2 \frac{\sigma^2}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}\end{aligned}$$

Further,

$$\begin{aligned}Cov(A, \varepsilon_i - \overline{\varepsilon}) &= Cov\left(\frac{\sum_{k=1}^n \varepsilon_k \left(X_k - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}, \varepsilon_i - \frac{\sum_k \varepsilon_k}{n}\right) \\ &= Cov\left(\frac{\sum_{k=1}^n \varepsilon_k \left(X_k - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}, \varepsilon_i\right) - Cov\left(\frac{\sum_{k=1}^n \varepsilon_k \left(X_k - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}, \frac{\sum_k \varepsilon_k}{n}\right) \\ &= \sigma^2 \frac{\left(X_i - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2} - \sum_{j=1}^n \frac{\sigma^2 \left(X_j - \overline{X}\right)}{n \sum_{k=1}^n \left(X_k - \overline{X}\right)^2} \\ &= \frac{\sigma^2 \left(X_i - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2} - \frac{\sigma^2 \sum_{j=1}^n \left(X_j - \overline{X}\right)}{n \sum_{k=1}^n \left(X_k - \overline{X}\right)^2} \\ &= \frac{\sigma^2 \left(X_i - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}, \text{ as } \sum_{j=1}^n \left(X_j - \overline{X}\right) = n\overline{X} - n\overline{X} = 0\end{aligned}$$

Therefore, (0.1) can be written as

$$\begin{aligned}Var(\widehat{\varepsilon}_i) &= \sigma^2\left(1 - \frac{1}{n}\right) + \left(X_i - \overline{X}\right)^2 \frac{\sigma^2}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2} - 2\left(X_i - \overline{X}\right) \frac{\sigma^2 \left(X_i - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2} \\ &= \sigma^2\left(1 - \frac{1}{n}\right) - \frac{\sigma^2 \left(X_i - \overline{X}\right)^2}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2} = \sigma^2\left(1 - \frac{1}{n} - \frac{\left(X_i - \overline{X}\right)^2}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}\right).\end{aligned}$$

(iv)

We assumed that ε_is are independent. But $\widehat{\varepsilon}_i$s can still be correlated. To see why $\{\widehat{\varepsilon}_i\}$ are correlated, consider, for example, the expressions for $\widehat{\varepsilon}_1$ and $\widehat{\varepsilon}_2$.

$$\begin{aligned}\widehat{\varepsilon}_1 &= \left(\varepsilon_1 - \frac{\sum_k \varepsilon_k}{n}\right) - \left(X_1 - \overline{X}\right) \frac{\sum_{k=1}^n \varepsilon_k \left(X_k - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2} \\ \widehat{\varepsilon}_2 &= \left(\varepsilon_2 - \frac{\sum_k \varepsilon_k}{n}\right) - \left(X_2 - \overline{X}\right) \frac{\sum_{k=1}^n \varepsilon_k \left(X_k - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}\end{aligned}$$

If $\widehat{\varepsilon}_1$ had been a function of ε_1 only, and $\widehat{\varepsilon}_2$ a function of ε_2, we would have zero covariance between $\widehat{\varepsilon}_1$ and $\widehat{\varepsilon}_2$. However, both $\widehat{\varepsilon}_1$ and $\widehat{\varepsilon}_2$ have all the disturbances terms $\{\varepsilon_i\}$ common, thus making them correlated. Moreover, the third term in $\widehat{\varepsilon}_i$ has all the error terms ε_k multiplied by $\frac{\left(X_i - \overline{X}\right)\left(X_k - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}$ respectively. This makes the covariance terms dependent on X_ks. Below we derive the covariance between $\widehat{\varepsilon}_i$ and $\widehat{\varepsilon}_j$.

(v)

$$\begin{aligned}(0.2) \qquad & Cov\left(\widehat{\varepsilon}_i, \widehat{\varepsilon}_j\right) \\ =\ & Cov\left(\varepsilon_i - \frac{\sum_k \varepsilon_k}{n} - \left(X_i - \overline{X}\right) \frac{\sum_{k=1}^n \varepsilon_k \left(X_k - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}, \varepsilon_j - \frac{\sum_k \varepsilon_k}{n} - \left(X_j - \overline{X}\right) \frac{\sum_{k=1}^n \varepsilon_k \left(X_k - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}\right) \\ =\ & Cov\left(\varepsilon_i - \frac{\sum_k \varepsilon_k}{n}, \varepsilon_j - \frac{\sum_k \varepsilon_k}{n}\right)' \\ & + Cov\left(\left(X_i - \overline{X}\right) \frac{\sum_{k=1}^n \varepsilon_k \left(X_k - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}, \left(X_j - \overline{X}\right) \frac{\sum_{k=1}^n \varepsilon_k \left(X_k - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}\right) \\ & - Cov\left(\varepsilon_i - \frac{\sum_k \varepsilon_k}{n}, \left(X_j - \overline{X}\right) \frac{\sum_{k=1}^n \varepsilon_k \left(X_k - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}\right) \\ & - Cov\left(\varepsilon_j - \frac{\sum_k \varepsilon_k}{n}, \left(X_i - \overline{X}\right) \frac{\sum_{k=1}^n \varepsilon_k \left(X_k - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}\right)\end{aligned}$$

The first term can be simplified as

$$\begin{aligned}& Cov\left(\varepsilon_i - \frac{\sum_k \varepsilon_k}{n}, \varepsilon_j - \frac{\sum_k \varepsilon_k}{n}\right) \\ & = Cov\left(\varepsilon_i, \varepsilon_j\right) + Cov\left(\frac{\sum_k \varepsilon_k}{n}, \frac{\sum_k \varepsilon_k}{n}\right) - Cov\left(\frac{\sum_k \varepsilon_k}{n}, \varepsilon_j\right) - Cov\left(\varepsilon_i, \frac{\sum_k \varepsilon_k}{n}\right) \\ & = 0 + \frac{n\sigma^2}{n^2} - \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = -\frac{\sigma^2}{n}\end{aligned}$$

The second term:

$$\begin{aligned}& Cov\left(\left(X_i - \overline{X}\right) \frac{\sum_{k=1}^n \varepsilon_k \left(X_k - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}, \left(X_j - \overline{X}\right) \frac{\sum_{k=1}^n \varepsilon_k \left(X_k - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}\right) \\ =\ & \frac{\left(X_i - \overline{X}\right)\left(X_j - \overline{X}\right)}{\left(\sum_{k=1}^n \left(X_k - \overline{X}\right)^2\right)^2} Cov\left(\sum_{k=1}^n \varepsilon_k \left(X_k - \overline{X}\right), \sum_{k=1}^n \varepsilon_k \left(X_k - \overline{X}\right)\right) \\ =\ & \frac{\left(X_i - \overline{X}\right)\left(X_j - \overline{X}\right)}{\left(\sum_{k=1}^n \left(X_k - \overline{X}\right)^2\right)^2} \sum_{k=1}^n \left(X_k - \overline{X}\right)^2 \sigma^2 \\ =\ & \frac{\left(X_i - \overline{X}\right)\left(X_j - \overline{X}\right)}{\left(\sum_{k=1}^n \left(X_k - \overline{X}\right)^2\right)} \sigma^2\end{aligned}$$

The third term:

$$
\begin{aligned}
& Cov\left(\varepsilon_i - \frac{\sum_k \varepsilon_k}{n}, \left(X_j - \overline{X}\right)\frac{\sum_{k=1}^n \varepsilon_k \left(X_k - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}\right) \\
= & Cov\left(\varepsilon_i, \left(X_j - \overline{X}\right)\frac{\sum_{k=1}^n \varepsilon_k \left(X_k - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}\right) - Cov\left(\frac{\sum_k \varepsilon_k}{n}, \left(X_j - \overline{X}\right)\frac{\sum_{k=1}^n \varepsilon_k \left(X_k - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}\right) \\
= & \frac{\left(X_j - \overline{X}\right)\left(X_i - \overline{X}\right) Cov\left(\varepsilon_i, \varepsilon_i\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2} - \frac{\left(X_j - \overline{X}\right)}{n\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}\sum_k \left(X_k - \overline{X}\right) Cov\left(\varepsilon_k, \varepsilon_k\right) \\
= & \frac{\left(X_j - \overline{X}\right)\left(X_i - \overline{X}\right)\sigma^2}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2} - \frac{\left(X_j - \overline{X}\right)\sigma^2}{n\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}\sum_k \left(X_k - \overline{X}\right) \\
& \text{Note that } \sum_k \left(X_k - \overline{X}\right) = 0, \\
= & \frac{\left(X_j - \overline{X}\right)\left(X_i - \overline{X}\right)\sigma^2}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}
\end{aligned}
$$

Similarly, it can be shown that

$$
Cov\left(\varepsilon_j - \frac{\sum_k \varepsilon_k}{n}, \left(X_i - \overline{X}\right)\frac{\sum_{k=1}^n \varepsilon_k \left(X_k - \overline{X}\right)}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}\right) = \frac{\left(X_i - \overline{X}\right)\left(X_j - \overline{X}\right)\sigma^2}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}.
$$

Therefore, (0.2) can be rewritten as

$$
\begin{aligned}
& Cov\left(\widehat{\varepsilon}_i, \widehat{\varepsilon}_j\right) \\
= & -\frac{\sigma^2}{n} + \frac{\left(X_i - \overline{X}\right)\left(X_j - \overline{X}\right)\sigma^2}{\left(\sum_{k=1}^n \left(X_k - \overline{X}\right)^2\right)} - 2\frac{\left(X_i - \overline{X}\right)\left(X_j - \overline{X}\right)}{\left(\sum_{k=1}^n \left(X_k - \overline{X}\right)^2\right)}\sigma^2 \\
= & -\sigma^2\left(\frac{1}{n} + \frac{\left(X_i - \overline{X}\right)\left(X_j - \overline{X}\right)}{\left(\sum_{k=1}^n \left(X_k - \overline{X}\right)^2\right)}\right)
\end{aligned}
$$

(vi)

Note that $E\left(\widehat{\varepsilon}_i^2\right) = Var\left(\widehat{\varepsilon}_i\right)$ as $E\left(\widehat{\varepsilon}_i\right) = 0$.

Therefore, we have

$$
\begin{aligned}
\sum_i E\left(\widehat{\varepsilon}_i^2\right) &= \sum_i Var\left(\widehat{\varepsilon}_i\right) \\
&= \sum_i \sigma^2\left(1 - \frac{1}{n} - \frac{\left(X_i - \overline{X}\right)^2}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}\right) \\
&= n\sigma^2\left(1 - \frac{1}{n}\right) - \frac{\sigma^2}{\sum_{k=1}^n \left(X_k - \overline{X}\right)^2}\sum_i \left(X_i - \overline{X}\right)^2 \\
&= n\sigma^2 - \sigma^2 - \sigma^2 = \sigma^2\left(n - 2\right).
\end{aligned}
$$