Proofs of Results in Chapter 7

Proof of Result 7.2 Before the response $\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ is observed, it is a random vector. Now,

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{\beta} + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon} \\
\hat{\boldsymbol{\varepsilon}} &= [\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{Y} \\
&= [\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'][\mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}] = [\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\boldsymbol{\varepsilon}
\end{aligned}$$

since $[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{Z} = \mathbf{Z} - \mathbf{Z} = \mathbf{0}$. From (2-24) and (2-45),

$$\begin{aligned}
E(\hat{\boldsymbol{\beta}}) &= \boldsymbol{\beta} + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'E(\boldsymbol{\varepsilon}) = \boldsymbol{\beta} \\
\operatorname{Cov}(\hat{\boldsymbol{\beta}}) &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\operatorname{Cov}(\boldsymbol{\varepsilon})\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} = \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \\
&= \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1} \\
E(\hat{\boldsymbol{\varepsilon}}) &= [\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']E(\boldsymbol{\varepsilon}) = \mathbf{0} \\
\operatorname{Cov}(\hat{\boldsymbol{\varepsilon}}) &= [\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\operatorname{Cov}(\boldsymbol{\varepsilon})[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']' \\
&= \sigma^2[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']
\end{aligned}$$

where the last equality follows from (7-6). Also,

$$\begin{aligned}
\operatorname{Cov}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\varepsilon}}) &= E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\hat{\boldsymbol{\varepsilon}}'] = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'] \\
&= \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'] = \mathbf{0}
\end{aligned}$$

because $\mathbf{Z}'[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'] = \mathbf{0}$. From the definition of $\hat{\boldsymbol{\varepsilon}}$ above, (7-6) and Result 4.9,

$$\begin{aligned}
\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} &= \boldsymbol{\varepsilon}'[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'][\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\boldsymbol{\varepsilon} \\
&= \boldsymbol{\varepsilon}'[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\boldsymbol{\varepsilon} \\
&= \operatorname{tr}[\boldsymbol{\varepsilon}'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\boldsymbol{\varepsilon}] \\
&= \operatorname{tr}([\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')
\end{aligned}$$

Now, for an arbitrary $n \times n$ random matrix $\mathbf{W}$,

$$\begin{aligned}
E(\operatorname{tr}(\mathbf{W})) &= E(W_{11} + W_{22} + \cdots + W_{nn}) \\
&= E(W_{11}) + E(W_{22}) + \cdots + E(W_{nn}) = \operatorname{tr}[E(\mathbf{W})]
\end{aligned}$$

Thus, using Result 2A.12, we obtain

$$\begin{aligned}
E(\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}) &= \operatorname{tr}([\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')) \\
&= \sigma^2 \operatorname{tr}[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'] \\
&= \sigma^2 \operatorname{tr}(\mathbf{I}) - \sigma^2 \operatorname{tr}[\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'] \\
&= \sigma^2 n - \sigma^2 \operatorname{tr}[(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Z}] \\
&= n\sigma^2 - \sigma^2 \operatorname{tr}\Big[\underset{(r+1)\times(r+1)}{\mathbf{I}}\Big] \\
&= \sigma^2(n - r - 1)
\end{aligned}$$

and the result for $s^2 = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}/(n - r - 1)$ follows. ∎

Proof of Result 7.4 Given the data and the normal assumption for the errors, the likelihood function for $\boldsymbol{\beta}, \sigma^2$ is

$$L(\boldsymbol{\beta}, \sigma^2) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\varepsilon_j^2/2\sigma^2} = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}/2\sigma^2}$$
$$= \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-(\mathbf{y}-\mathbf{Z}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{Z}\boldsymbol{\beta})/2\sigma^2}$$

For a fixed value σ^2, the likelihood is maximized by minimizing $(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})$. But this minimization yields the least squares estimate $\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$, which does not depend upon σ^2. Therefore, under the normal assumption, the maximum likelihood and least squares approaches provide the same estimator $\hat{\boldsymbol{\beta}}$. Next, maximizing $L(\hat{\boldsymbol{\beta}}, \sigma^2)$ over σ^2 [see (4-18)] gives

$$L(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) = \frac{1}{(2\pi)^{n/2}(\hat{\sigma}^2)^{n/2}} e^{-n/2} \quad \text{where} \quad \hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\beta}})}{n}$$

As shown in the proof of Result 7.2, we can express $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\varepsilon}}$ as linear combinations of the normal variables $\boldsymbol{\varepsilon}$. Specifically,

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hline \hat{\boldsymbol{\varepsilon}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta} + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon} \\ \hline [\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta} \\ \hline \mathbf{0} \end{bmatrix} + \begin{bmatrix} (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' \\ \hline \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' \end{bmatrix} \boldsymbol{\varepsilon} = \boldsymbol{\alpha} + \mathbf{A}\boldsymbol{\varepsilon}$$

Because $\mathbf{Z}$ is fixed, Result 4.3 implies the joint normality of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\varepsilon}}$. Their mean vectors and covariance matrices were obtained in Result 7.2. Again, using (7-6), we get

$$\text{Cov}\left(\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hline \hat{\boldsymbol{\varepsilon}} \end{bmatrix}\right) = \mathbf{A}\,\text{Cov}(\boldsymbol{\varepsilon})\mathbf{A}' = \sigma^2 \left[\begin{array}{c|c} (\mathbf{Z}'\mathbf{Z})^{-1} & \mathbf{0} \\ \hline \mathbf{0}' & \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' \end{array}\right]$$

Since $\text{Cov}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\varepsilon}}) = \mathbf{0}$ for the normal random vectors $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\varepsilon}}$, these vectors are independent. (See Result 4.5.)

Next, let $(\lambda, \mathbf{e})$ be any eigenvalue–eigenvector pair for $\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$. Then, by (7-6), $[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']^2 = [\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']$ so

$$\lambda\mathbf{e} = [\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{e} = [\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']^2\mathbf{e} = \lambda[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{e} = \lambda^2\mathbf{e}$$

That is, $\lambda = 0$ or 1. Now, $\text{tr}[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'] = n - r - 1$ (see the proof of Result 7.2), and from Result 4.9, $\text{tr}[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'] = \lambda_1 + \lambda_2 + \cdots + \lambda_n$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are the eigenvalues of $[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'.]$ Consequently, exactly $n - r - 1$ values of λ_i equal one, and the rest are zero. It then follows from the spectral decomposition that

$$[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'] = \mathbf{e}_1\mathbf{e}_1' + \mathbf{e}_2\mathbf{e}_2' + \cdots + \mathbf{e}_{n-r-1}\mathbf{e}_{n-r-1}'$$

where $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{n-r-1}$ are the normalized eigenvectors associated with the eigenvalues $\lambda_1 = \lambda_2 = \cdots = \lambda_{n-r-1} = 1$. Let

$$\mathbf{V} = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_{n-r-1} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1' \\ \hline \mathbf{e}_2' \\ \hline \vdots \\ \hline \mathbf{e}_{n-r-1}' \end{bmatrix} \boldsymbol{\varepsilon}$$

Then $\mathbf{V}$ is normal with mean vector $\mathbf{0}$ and

$$\text{Cov}(V_i, V_k) = \begin{cases} \mathbf{e}_i'\sigma^2\mathbf{I}\mathbf{e}_k = \sigma^2\mathbf{e}_i'\mathbf{e}_k = \sigma^2, & i = k \\ 0, & \text{otherwise} \end{cases}$$

That is, the V_i are independent $N(0, \sigma^2)$ and

$$n\hat{\sigma}^2 = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon}'[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\boldsymbol{\varepsilon} = V_1^2 + V_2^2 + \cdots + V_{n-r-1}^2$$

is distributed $\sigma^2\chi^2_{n-r-1}$. ■

Proof of Result 7.10 According to the regression model, the likelihood is determined from the data $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n]'$ whose rows are independent, with $\mathbf{Y}_j$ distributed as $N_m(\boldsymbol{\beta}'\mathbf{z}_j, \boldsymbol{\Sigma})$. We first note that $\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta} = [\mathbf{Y}_1 - \boldsymbol{\beta}'\mathbf{z}_1, \mathbf{Y}_2 - \boldsymbol{\beta}'\mathbf{z}_2, \ldots, \mathbf{Y}_n - \boldsymbol{\beta}'\mathbf{z}_n]'$ so

$$(\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta}) = \sum_{j=1}^{n} (\mathbf{Y}_j - \boldsymbol{\beta}'\mathbf{z}_j)(\mathbf{Y}_j - \boldsymbol{\beta}'\mathbf{z}_j)'$$

and

$$\begin{aligned}\sum_{j=1}^{n} (\mathbf{Y}_j - \boldsymbol{\beta}'\mathbf{z}_j)'\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_j - \boldsymbol{\beta}'\mathbf{z}_j) &= \sum_{j=1}^{n} \operatorname{tr}[(\mathbf{Y}_j - \boldsymbol{\beta}'\mathbf{z}_j)'\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_j - \boldsymbol{\beta}'\mathbf{z}_j)] \\ &= \sum_{j=1}^{n} \operatorname{tr}[\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_j - \boldsymbol{\beta}'\mathbf{z}_j)(\mathbf{Y}_j - \boldsymbol{\beta}'\mathbf{z}_j)'] \\ &= \operatorname{tr}[\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta})]\end{aligned}$$

Another preliminary calculation will enable us to express the likelihood in a simple form. Since $\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}$ satisfies $\mathbf{Z}'\hat{\boldsymbol{\varepsilon}} = \mathbf{0}$ [(see 7-29)],

$$\begin{aligned}&(\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta}) \\ &\quad = [\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}} + \mathbf{Z}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})]'[\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}} + \mathbf{Z}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})] \\ &\quad = (\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{Z}'\mathbf{Z}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ &\quad = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{Z}'\mathbf{Z}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\end{aligned}$$

Using the results above, we obtain the likelihood

$$\begin{aligned}L(\boldsymbol{\beta}, \boldsymbol{\Sigma}) &= \prod_{j=1}^{n} \frac{1}{(2\pi)^{m/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{y}_j - \boldsymbol{\beta}'\mathbf{z}_j)'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_j - \boldsymbol{\beta}'\mathbf{z}_j)} \\ &= \frac{1}{(2\pi)^{mn/2}} \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{1}{2}\operatorname{tr}[\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{Z}'\mathbf{Z}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}))]} \\ &= \frac{1}{(2\pi)^{mn/2}} \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{1}{2}\operatorname{tr}[\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}] - \frac{1}{2}\operatorname{tr}[\mathbf{Z}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{Z}']}\end{aligned}$$

The matrix $\mathbf{Z}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{Z}'$ is the form $\mathbf{A}'\mathbf{A}$, with $\mathbf{A} = \boldsymbol{\Sigma}^{-1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{Z}'$, and, from Exercise 2.16, it is nonnegative definite. Therefore, its eigenvalues are nonnegative also. Since, by Result 4.9, $\operatorname{tr}[\mathbf{Z}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{Z}']$ is the sum of its eigenvalues, this trace will equal its minimum value, zero, if $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$. This choice is unique because $\mathbf{Z}$ is of full rank and $\hat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}_{(i)} \neq \mathbf{0}$, implies that $\mathbf{Z}(\hat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}_{(i)}) \neq \mathbf{0}$, in which case $\operatorname{tr}[\mathbf{Z}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{Z}'] \geq \mathbf{c}'\boldsymbol{\Sigma}^{-1}\mathbf{c} > 0$, where $\mathbf{c}'$ is any nonzero row of $\mathbf{Z}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. Applying Result 4.10 with $\mathbf{B} = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$, $b = n/2$, and $p = m$, we find that $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Sigma}} = n^{-1}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$ are the maximum likelihood estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, respectively, and

$$L(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}}) = \frac{1}{(2\pi)^{mn/2}} \frac{(n)^{mn/2}}{|\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}|^{n/2}} e^{-nm/2} = \frac{e^{-nm/2}}{(2\pi)^{mn/2}|\hat{\boldsymbol{\Sigma}}|^{n/2}}$$

It remains to establish the distributional results. From (7-33), we know that $\hat{\boldsymbol{\beta}}_{(i)}$ and $\hat{\boldsymbol{\varepsilon}}_{(i)}$ are linear combinations of the elements of $\boldsymbol{\varepsilon}$. Specifically,

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{(i)} &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}_{(i)} + \boldsymbol{\beta}_{(i)} \\ \hat{\boldsymbol{\varepsilon}}_{(i)} &= [\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\boldsymbol{\varepsilon}_{(i)}, \qquad i = 1, 2, \ldots, m\end{aligned}$$

Therefore, by Result 4.3, $\hat{\boldsymbol{\beta}}_{(1)}, \hat{\boldsymbol{\beta}}_{(2)}, \ldots, \hat{\boldsymbol{\beta}}_{(m)}, \hat{\boldsymbol{\varepsilon}}_{(1)}, \hat{\boldsymbol{\varepsilon}}_{(2)}, \ldots, \hat{\boldsymbol{\varepsilon}}_{(m)}$ are jointly normal. Their mean vectors and covariance matrices are given in Result 7.9. Since $\hat{\boldsymbol{\varepsilon}}$ and $\hat{\boldsymbol{\beta}}$ have a zero covariance matrix, by Result 4.5 they are independent. From the proof of Result 7.4, $[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'] = \sum_{\ell=1}^{n-r-1} \mathbf{e}_\ell \mathbf{e}'_\ell$, where $\mathbf{e}'_\ell \mathbf{e}_k = 0$, $\ell \neq k$, and $\mathbf{e}'_\ell \mathbf{e}_\ell = 1$. Set $\mathbf{V}_\ell = \boldsymbol{\varepsilon}'\mathbf{e}_\ell = [\boldsymbol{\varepsilon}'_{(1)}\mathbf{e}_\ell, \boldsymbol{\varepsilon}'_{(2)}\mathbf{e}_\ell, \ldots, \boldsymbol{\varepsilon}'_{(m)}\mathbf{e}_\ell]' = e_{\ell 1}\boldsymbol{\varepsilon}_1 + e_{\ell 2}\boldsymbol{\varepsilon}_2 + \cdots + e_{\ell n}\boldsymbol{\varepsilon}_n$. Because $\mathbf{V}_\ell$, $\ell = 1, 2, \ldots, n-r-1$, are linear combinations of the elements of $\boldsymbol{\varepsilon}$, they have a joint normal distribution with $E(\mathbf{V}_\ell) = E(\boldsymbol{\varepsilon}')\mathbf{e}_\ell = \mathbf{0}$. Also, by Result 4.8, $\mathbf{V}_\ell$ and $\mathbf{V}_k$ have covariance matrix $(\mathbf{e}'_\ell \mathbf{e}_k)\boldsymbol{\Sigma} = (0)\boldsymbol{\Sigma} = \mathbf{0}$ if $\ell \neq k$. Consequently, the $\mathbf{V}_\ell$ are independently distributed as $N_m(\mathbf{0}, \boldsymbol{\Sigma})$. Finally,

$$\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon}'[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\boldsymbol{\varepsilon} = \sum_{\ell=1}^{n-r-1} \boldsymbol{\varepsilon}'\mathbf{e}_\ell \mathbf{e}'_\ell \boldsymbol{\varepsilon} = \sum_{\ell=1}^{n-r-1} \mathbf{V}_\ell \mathbf{V}'_\ell$$

which has the $W_{p,n-r-1}(\boldsymbol{\Sigma})$ distribution, by (4-22). ∎