same answer, if it is feasible to calculate $\mathfrak{I}(\boldsymbol{\theta})$ at all, although one approach may be easier than the other in any given situation.

For the nonlinear regression model (8.79), the parameter vector $\boldsymbol{\theta}$ is the vector $[\boldsymbol{\beta} \,\vdots\, \sigma]$. We now calculate the limiting information matrix $\mathfrak{I}(\boldsymbol{\beta}, \sigma)$ for this model using the second method, based on the CG matrix, which requires only first derivatives. It is a good exercise to repeat the derivation using the Hessian, which requires second derivatives, and verify that it yields the same results. The first derivative of $\ell_t(y_t, \boldsymbol{\beta}, \sigma)$ with respect to β_i is

$$\frac{\partial \ell_t}{\partial \beta_i} = \frac{1}{\sigma^2}\big(y_t - x_t(\boldsymbol{\beta})\big)X_{ti}(\boldsymbol{\beta}) = \frac{1}{\sigma^2}\, e_t(\boldsymbol{\beta})X_{ti}(\boldsymbol{\beta}), \tag{8.83}$$

where $e_t(\boldsymbol{\beta}) \equiv y_t - x_t(\boldsymbol{\beta})$ and, as usual, $X_{ti}(\boldsymbol{\beta}) \equiv \partial x_t(\boldsymbol{\beta})/\partial \beta_i$. The first derivative of $\ell_t(y_t, \boldsymbol{\beta}, \sigma)$ with respect to σ is

$$\frac{\partial \ell_t}{\partial \sigma} = -\frac{1}{\sigma} + \frac{\big(y_t - x_t(\boldsymbol{\beta})\big)^2}{\sigma^3} = -\frac{1}{\sigma} + \frac{e_t^2(\boldsymbol{\beta})}{\sigma^3}. \tag{8.84}$$

Expressions (8.83) and (8.84) are all that we need to calculate the information matrix using the CG matrix. The column of that matrix which corresponds to σ will have typical element (8.84), while the remaining k columns, which correspond to the β_i's, will have typical element (8.83).

The element of $\mathfrak{I}(\boldsymbol{\beta}, \sigma)$ corresponding to β_i and β_j is

$$\mathfrak{I}(\beta_i, \beta_j) = \operatorname*{plim}_{n\to\infty}\bigg(\frac{1}{n}\sum_{t=1}^{n}\frac{e_t^2(\boldsymbol{\beta})}{\sigma^4}X_{ti}(\boldsymbol{\beta})X_{tj}(\boldsymbol{\beta})\bigg).$$

Since $e_t^2(\boldsymbol{\beta})$ has expectation σ^2 under the DGP characterized by $(\boldsymbol{\beta}, \sigma)$ and is independent of $\boldsymbol{X}(\boldsymbol{\beta})$, we can replace it by σ^2 here to yield

$$\mathfrak{I}(\beta_i, \beta_j) = \operatorname*{plim}_{n\to\infty}\bigg(\frac{1}{n}\sum_{t=1}^{n}\frac{1}{\sigma^2}X_{ti}(\boldsymbol{\beta})X_{tj}(\boldsymbol{\beta})\bigg).$$

Thus we see that the whole $(\boldsymbol{\beta}, \boldsymbol{\beta})$ block of the limiting information matrix is

$$\frac{1}{\sigma^2}\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\boldsymbol{X}^{\top}(\boldsymbol{\beta})\boldsymbol{X}(\boldsymbol{\beta})\Big). \tag{8.85}$$

The element of $\mathfrak{I}(\boldsymbol{\beta}, \sigma)$ corresponding to σ is

$$\begin{aligned}
\mathfrak{I}(\sigma, \sigma) &= \operatorname*{plim}_{n\to\infty}\bigg(\frac{1}{n}\sum_{t=1}^{n}\bigg(\frac{1}{\sigma^2} + \frac{e_t^4(\boldsymbol{\beta})}{\sigma^6} - \frac{2e_t^2(\boldsymbol{\beta})}{\sigma^4}\bigg)\bigg) \\
&= \frac{1}{n}\bigg(\frac{n}{\sigma^2} + \frac{3n\sigma^4}{\sigma^6} - \frac{2n\sigma^2}{\sigma^4}\bigg) \\
&= \frac{2}{\sigma^2}.
\end{aligned} \tag{8.86}$$