Next, we derive a useful and general result that will allow us to replace the vector of derivatives $\boldsymbol{\mu}_0$ in (16.54) by something more manageable. The moment condition under test is given by (16.48). The moment can be written out explicitly as

$$E_{\boldsymbol{\theta}}\big(m_t(y_t,\boldsymbol{\theta})\big) = \int_{-\infty}^{\infty} m_t(y_t,\boldsymbol{\theta})L_t(y_t,\boldsymbol{\theta})dy_t. \tag{16.55}$$

Differentiating the right-hand side of (16.55) with respect to the components of $\boldsymbol{\theta}$, we obtain, by the same sort of reasoning as led to the information matrix equality (8.44),

$$E_{\boldsymbol{\theta}}\big(m_t(\boldsymbol{\theta})\boldsymbol{G}_t(\boldsymbol{\theta})\big) = -E_{\boldsymbol{\theta}}\big(\boldsymbol{N}_t(\boldsymbol{\theta})\big). \tag{16.56}$$

Here $\boldsymbol{G}_t(\boldsymbol{\theta})$ is the contribution made by observation $t$ to the gradient of the loglikelihood function, and the $1 \times k$ row vector $\boldsymbol{N}_t(\boldsymbol{\theta})$ has typical element $\partial m_t(\boldsymbol{\theta})/\partial\theta_i$.[5] The most useful form of our result is obtained by summing (16.56) over $t$. Let $\boldsymbol{m}(\boldsymbol{\theta})$ be an $n$–vector with typical element $m_t(\boldsymbol{\theta})$, and let $\boldsymbol{N}(\boldsymbol{\theta})$ be an $n \times k$ matrix with typical row $\boldsymbol{N}_t(\boldsymbol{\theta})$. Then

$$\frac{1}{n}E_{\boldsymbol{\theta}}\big(\boldsymbol{G}^\top(\boldsymbol{\theta})\boldsymbol{m}(\boldsymbol{\theta})\big) = -\frac{1}{n}E_{\boldsymbol{\theta}}\big(\boldsymbol{N}^\top(\boldsymbol{\theta})\boldsymbol{\iota}\big), \tag{16.57}$$

where, as usual, $\boldsymbol{G}(\boldsymbol{\theta})$ denotes the CG matrix. In (16.54), $\boldsymbol{\mu}_0 = n^{-1}\boldsymbol{N}_0^\top\boldsymbol{\iota}$, where $\boldsymbol{N}_0 \equiv \boldsymbol{N}(\boldsymbol{\theta}_0)$. By the law of large numbers, this will converge to the limit of the right-hand side of (16.57), and so also to the limit of the left-hand side. Thus, if $\boldsymbol{G}_0 \equiv \boldsymbol{G}(\boldsymbol{\theta}_0)$, we can assert that

$$\boldsymbol{\mu}_0 = \frac{1}{n}\boldsymbol{N}_0^\top\boldsymbol{\iota} \overset{a}{=} -\frac{1}{n}\boldsymbol{G}_0^\top\boldsymbol{m}_0. \tag{16.58}$$

We next make use of the very well-known result (13.18) on the relationship between ML estimates, the information matrix, and the score vector:

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \overset{a}{=} \mathcal{I}_0^{-1}n^{-1/2}\boldsymbol{g}_0. \tag{16.59}$$

Since the information matrix $\mathcal{I}_0$ is asymptotically equal to $n^{-1}\boldsymbol{G}_0^\top\boldsymbol{G}_0$ (see Section 8.6), and $\boldsymbol{g}_0 = \boldsymbol{G}_0^\top\boldsymbol{\iota}$, (16.59) becomes

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \overset{a}{=} \big(n^{-1}\boldsymbol{G}_0^\top\boldsymbol{G}_0\big)^{-1}n^{-1/2}\boldsymbol{G}_0^\top\boldsymbol{\iota}.$$

This result, combined with (16.58), allows us to replace the right-hand side of (16.54) by

$$n^{-1/2}\boldsymbol{m}_0^\top\boldsymbol{\iota} - n^{-1}\boldsymbol{m}_0^\top\boldsymbol{G}_0\big(n^{-1}\boldsymbol{G}_0^\top\boldsymbol{G}_0\big)^{-1}n^{-1/2}\boldsymbol{G}_0^\top\boldsymbol{\iota} = n^{-1/2}\boldsymbol{m}_0^\top\boldsymbol{M}_G\boldsymbol{\iota}, \tag{16.60}$$

where $\boldsymbol{M}_G$ denotes the matrix that projects orthogonally onto $\mathcal{S}^\perp(\boldsymbol{G}_0)$.

[5] Our usual notation would have been $\boldsymbol{M}_t(\boldsymbol{\theta})$ instead of $\boldsymbol{N}_t(\boldsymbol{\theta})$, but this would conflict with the standard notation for complementary orthogonal projections.