as can be seen directly from (13.82). Since the asymptotic equivalence of $\grave{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}$ requires the factors of $n^{1/2}$ that appear in (13.83), it can be seen why we wish to prove (13.82), with a factor of $n^{1/2}$ on each side of the equation, rather than the seemingly equivalent result that $\acute{\boldsymbol{c}} \stackrel{a}{=} \hat{\boldsymbol{\theta}} - \acute{\boldsymbol{\theta}}$. Although this result is certainly true, it is weaker than (13.82), because it merely implies that $\hat{\boldsymbol{\theta}} - \acute{\boldsymbol{\theta}} = o(1)$, while (13.82) implies that $\hat{\boldsymbol{\theta}} - \acute{\boldsymbol{\theta}} = o(n^{-1/2})$.

The proof of (13.82) is both simple and illuminating. A Taylor expansion of the gradient $\acute{\boldsymbol{g}} \equiv \boldsymbol{g}(\acute{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}_0$ yields

$$n^{-1/2}\acute{\boldsymbol{g}} = n^{-1/2}\boldsymbol{g}_0 + n^{-1}\boldsymbol{H}(\boldsymbol{\theta}_0)n^{1/2}(\acute{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + O(n^{-1/2}),$$

where, as usual, $\boldsymbol{H}(\boldsymbol{\theta})$ denotes the Hessian of the loglikelihood function $\ell(\boldsymbol{\theta})$. If now we expand $\hat{\boldsymbol{g}}$, which is zero by the first-order conditions for a maximum of the likelihood at $\hat{\boldsymbol{\theta}}$, we obtain

$$\mathbf{0} = n^{-1/2}\boldsymbol{g}_0 + n^{-1}\boldsymbol{H}(\boldsymbol{\theta}_0)n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + O(n^{-1/2}).$$

On subtracting the last two equations and noting that $\acute{\boldsymbol{g}} = \acute{\boldsymbol{G}}^{\top}\boldsymbol{\iota}$, we find that

$$n^{-1/2}\acute{\boldsymbol{G}}^{\top}\boldsymbol{\iota} = n^{-1}\boldsymbol{H}(\boldsymbol{\theta}_0)n^{1/2}(\acute{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) + O(n^{-1/2}). \tag{13.84}$$

By the information matrix equality, $n^{-1}\boldsymbol{H}(\boldsymbol{\theta}_0) = -\mathcal{I}_0 + o(1)$. Since, by the consistency of $\acute{\boldsymbol{\theta}}$, we have $n^{-1}\acute{\boldsymbol{G}}^{\top}\acute{\boldsymbol{G}} = \mathcal{I}_0 + o(1)$, we may replace $n^{-1}\boldsymbol{H}(\boldsymbol{\theta}_0)$ in (13.84) by $-n^{-1}\acute{\boldsymbol{G}}^{\top}\acute{\boldsymbol{G}}$ to obtain

$$n^{-1/2}\acute{\boldsymbol{G}}^{\top}\boldsymbol{\iota} = \big(n^{-1}\acute{\boldsymbol{G}}^{\top}\acute{\boldsymbol{G}}\big)n^{1/2}(\hat{\boldsymbol{\theta}} - \acute{\boldsymbol{\theta}}) + o(1).$$

The result (13.82) now follows directly on premultiplication by $(n^{-1}\acute{\boldsymbol{G}}^{\top}\acute{\boldsymbol{G}})^{-1}$.

A second property of artificial regressions is the one that permits their use in the calculation of LM statistics. When an artificial regression that satisfies this property is evaluated at a root-n consistent $\acute{\boldsymbol{\theta}}$, n times the uncentered R^2 calculated from it is asymptotically equal to

$$\tfrac{1}{n}\acute{\boldsymbol{g}}^{\top}\mathcal{I}_0^{-1}\acute{\boldsymbol{g}}.$$

This result is very easy to prove for the OPG regression. The R^2 is the ratio of the explained sum of squares (ESS) to the total sum of squares (TSS), and so nR^2 is the ratio ESS/(TSS/n). We saw that TSS/n was equal to 1. This means that nR^2 is just the explained sum of squares:

$$nR^2 = \boldsymbol{\iota}^{\top}\acute{\boldsymbol{G}}\big(\acute{\boldsymbol{G}}^{\top}\acute{\boldsymbol{G}}\big)^{-1}\acute{\boldsymbol{G}}^{\top}\boldsymbol{\iota} = \tfrac{1}{n}\acute{\boldsymbol{g}}^{\top}\big(n^{-1}\acute{\boldsymbol{G}}^{\top}\acute{\boldsymbol{G}}\big)^{-1}\acute{\boldsymbol{g}}. \tag{13.85}$$

This completes the proof, since $n^{-1}\acute{\boldsymbol{G}}^{\top}\acute{\boldsymbol{G}} \to \mathcal{I}_0$.