Luckily, alternative methods of calculating IM tests are available in many cases. These invariably have much better finite-sample properties than the OPG version but are not as widely applicable. Various techniques have been suggested by Chesher and Spady (1991), Orme (1990a, 1990b), and Davidson and MacKinnon (1992a). In the last of these papers, we made use of an important result due to Chesher (1984), who showed that the implicit alternative of the IM test is a model with random parameter variation. This allowed us explicitly to construct a test against this type of alternative for the class of models to which the DLR is applicable (see Section 14.4). Orme (1990b) suggests alternative varieties of double- and even triple-length regressions for computing IM tests in other types of models.

Obtaining an IM test statistic that is inconsistent with the null hypothesis (which might have to be a very big number indeed if the OPG version of the test is being used), does not necessarily mean that one has to abandon the model being tested. What it does mean is that one has to use more robust methods of inference. In the case of regression models, we saw in Section 16.3 that one can make valid inferences in the presence of heteroskedasticity of unknown form by using an HCCME instead of the conventional least squares covariance matrix. In the more general case of models estimated by maximum likelihood, a similar option is open to us. Recall the result

$$\boldsymbol{V}^{\infty}\big(n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\big) = \mathcal{H}^{-1}(\boldsymbol{\theta}_0)\mathcal{I}(\boldsymbol{\theta}_0)\mathcal{H}^{-1}(\boldsymbol{\theta}_0), \tag{16.73}$$

which was originally (8.42). We obtained this result before we proved the information matrix equality, which we used to obtain the simpler result that

$$\boldsymbol{V}^{\infty}\big(n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\big) = \mathcal{I}^{-1}(\boldsymbol{\theta}_0). \tag{16.74}$$

Moreover, the assumptions used to obtain (16.73) were not as strong as those used to obtain the information matrix equality. This suggests that (16.73) may be true more generally than (16.74), and that is indeed the case, as White (1982) has shown. Thus, if there is reason to believe that the information matrix equality does not hold, it may be a good idea to employ the following estimator for the covariance matrix of $\hat{\boldsymbol{\theta}}$:

$$\hat{\boldsymbol{H}}^{-1}(\hat{\boldsymbol{G}}^{\top}\hat{\boldsymbol{G}})\hat{\boldsymbol{H}}^{-1}, \tag{16.75}$$

where $\hat{\boldsymbol{H}}$ denotes the Hessian matrix evaluated at the ML estimates $\hat{\boldsymbol{\theta}}$. As the natural analog of (8.42), expression (16.75) will be asymptotically valid under weaker conditions than either $-\hat{\boldsymbol{H}}^{-1}$ or $(\hat{\boldsymbol{G}}^{\top}\hat{\boldsymbol{G}})^{-1}$.