The advantage of this estimator is that it normally involves fewer random terms than does the empirical Hessian, and it may therefore be somewhat more efficient in finite samples. In the case of the classical normal linear model, to be discussed below, it is not at all difficult to obtain $\boldsymbol{I}(\boldsymbol{\theta})$, and the information matrix estimator is therefore the one that is normally used.

The third method is based on (10.31), from which we see that

$$\boldsymbol{I}(\boldsymbol{\theta}_0) = \mathrm{E}\big(\boldsymbol{G}^\top(\boldsymbol{\theta}_0)\boldsymbol{G}(\boldsymbol{\theta}_0)\big).$$

We can therefore estimate $n^{-1}\boldsymbol{I}(\boldsymbol{\theta}_0)$ consistently by $n^{-1}\boldsymbol{G}^\top(\hat{\boldsymbol{\theta}})\boldsymbol{G}(\hat{\boldsymbol{\theta}})$. The corresponding estimator of the covariance matrix, which is usually called the **outer-product-of-the-gradient**, or **OPG**, estimator, is

$$\widehat{\mathrm{Var}}_{\mathrm{OPG}}(\hat{\boldsymbol{\theta}}) = \big(\boldsymbol{G}^\top(\hat{\boldsymbol{\theta}})\boldsymbol{G}(\hat{\boldsymbol{\theta}})\big)^{-1}. \tag{10.44}$$

The OPG estimator has the advantage of being very easy to calculate. Unlike the empirical Hessian, it depends solely on first derivatives. Unlike the IM estimator, it requires no theoretical calculations. However, it tends to be less reliable in finite samples than either of the other two. The OPG estimator is sometimes called the **BHHH** estimator, because it was advocated by Berndt, Hall, Hall, and Hausman (1974) in a very well-known paper.

In practice, the estimators (10.42), (10.43), and (10.44) are all commonly used to estimate the covariance matrix of ML estimates, but many other estimators are available for particular models. Often, it may be difficult to obtain $\boldsymbol{I}(\boldsymbol{\theta})$, but not difficult to obtain another matrix that approximates it asymptotically, by starting either from the matrix $-\boldsymbol{H}(\boldsymbol{\theta})$ or from the matrix $\boldsymbol{G}^\top(\boldsymbol{\theta})\boldsymbol{G}(\boldsymbol{\theta})$ and taking expectations of some elements.

A fourth covariance matrix estimator, which follows directly from (10.40), is the **sandwich estimator**

$$\widehat{\mathrm{Var}}_{\mathrm{S}}(\hat{\boldsymbol{\theta}}) = \boldsymbol{H}^{-1}(\hat{\boldsymbol{\theta}})\boldsymbol{G}^\top(\hat{\boldsymbol{\theta}})\boldsymbol{G}(\hat{\boldsymbol{\theta}})\boldsymbol{H}^{-1}(\hat{\boldsymbol{\theta}}). \tag{10.45}$$

In normal circumstances, this estimator has little to recommend it. It is harder to compute than the OPG estimator and can be just as unreliable in finite samples. However, unlike the other three estimators, it is valid even when the information matrix equality does not hold. Since this equality generally fails to hold when the model is misspecified, it may be desirable to compute (10.45) and compare it with the other estimators.

When an ML estimator is applied to a model which is misspecified in ways that do not affect the consistency of the estimator, it is said to be a **quasi-ML estimator**, or **QMLE**; see White (1982) and Gouriéroux, Monfort, and Trognon (1984). In general, the sandwich covariance matrix estimator (10.45) is valid for QML estimators, but the other covariance matrix estimators, which depend on the information matrix equality, are not valid. At least, they are