The LM statistic (8.76) is numerically equal to a test based on the score vector $\boldsymbol{g}(\tilde{\boldsymbol{\theta}})$. By the first set of first-order conditions (8.72), $\boldsymbol{g}(\tilde{\boldsymbol{\theta}}) = \tilde{\boldsymbol{R}}^\top\tilde{\boldsymbol{\lambda}}$. Substituting $\boldsymbol{g}(\tilde{\boldsymbol{\theta}})$ for $\tilde{\boldsymbol{R}}^\top\tilde{\boldsymbol{\lambda}}$ in (8.76) yields the score form of the LM test,

$$\frac{1}{n}\tilde{\boldsymbol{g}}^\top\tilde{\mathfrak{I}}^{-1}\tilde{\boldsymbol{g}}. \tag{8.77}$$

In practice, this score form is often more useful than the LM form because, since restricted estimates are rarely obtained via a Lagrangian, $\tilde{\boldsymbol{g}}$ is generally readily available while $\tilde{\boldsymbol{\lambda}}$ typically is not. However, deriving the test via the Lagrange multipliers is illuminating, because this derivation makes it quite clear why the test has r degrees of freedom.

The third of the three classical tests is the **Wald test**. This test is very easy to derive. It asks whether the vector of restrictions, evaluated at the unrestricted estimates, is close enough to a zero vector for the restrictions to be plausible. In the case of the restrictions (8.71), the Wald test is based on the vector $\boldsymbol{r}(\hat{\boldsymbol{\theta}})$, which should tend to a zero vector asymptotically if the restrictions hold. As we have seen in Sections 8.5 and 8.6,

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \overset{a}{\sim} N\big(\mathbf{0},\, \mathfrak{I}^{-1}(\boldsymbol{\theta}_0)\big).$$

A Taylor-series approximation of $\boldsymbol{r}(\hat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}_0$ yields $\boldsymbol{r}(\hat{\boldsymbol{\theta}}) \cong \boldsymbol{R}_0(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$. Therefore,

$$\boldsymbol{V}\big(n^{1/2}\boldsymbol{r}(\hat{\boldsymbol{\theta}})\big) \overset{a}{=} \boldsymbol{R}_0\,\mathfrak{I}_0^{-1}\boldsymbol{R}_0^\top.$$

It follows that an appropriate test statistic is

$$n\boldsymbol{r}^\top(\hat{\boldsymbol{\theta}})\big(\hat{\boldsymbol{R}}\hat{\mathfrak{I}}^{-1}\hat{\boldsymbol{R}}^\top\big)^{-1}\boldsymbol{r}(\hat{\boldsymbol{\theta}}), \tag{8.78}$$

where $\hat{\mathfrak{I}}$ denotes any consistent estimate of $\mathfrak{I}(\boldsymbol{\theta}_0)$ based on the unrestricted estimates $\hat{\boldsymbol{\theta}}$. Different variants of the Wald test will use different estimates of $\mathfrak{I}(\boldsymbol{\theta}_0)$. It is easy to see that given suitable regularity the test statistic (8.78) will be asymptotically distributed as $\chi^2(r)$ under the null.

The fundamental property of the three classical test statistics is that under the null hypothesis, as $n \to \infty$, they all tend to the same random variable, which is distributed as $\chi^2(r)$. We will prove this result in Chapter 13. The implication is that, in large samples, it does not really matter which of the three tests we use. If both $\hat{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}$ are easy to compute, it is attractive to use the LR test. If $\tilde{\boldsymbol{\theta}}$ is easy to compute but $\hat{\boldsymbol{\theta}}$ is not, as is often the case for tests of model specification, then the LM test becomes attractive. If on the other hand $\hat{\boldsymbol{\theta}}$ is easy to compute but $\tilde{\boldsymbol{\theta}}$ is not, as may be the case when we are interested in nonlinear restrictions on a linear model, then the Wald test becomes attractive. When the sample size is not large, choice among the three tests is complicated by the fact that they may have very different finite-sample properties, which may further differ greatly among the alternative variants of the LM and Wald tests. This makes the choice of tests rather more complicated in practice than asymptotic theory would suggest.