the other two, and they can sometimes overreject severely. The performance of alternative Wald tests in models like (10.61) has been investigated by Gregory and Veall (1985, 1987). Other cases in which Wald tests perform very badly are discussed by Lafontaine and White (1986).

Because of their dubious finite-sample properties and their sensitivity to the way in which the restrictions are written, we recommend against using Wald tests when the outcome of a test is important, except when it would be very costly or inconvenient to estimate the restricted model. Asymptotic t statistics should also be used with great caution, since, as we saw in Section 6.7, every asymptotic t statistic is simply the signed square root of a Wald statistic. Because conventional confidence intervals are based on inverting asymptotic t statistics, they too should be used with caution.

Lagrange Multiplier Tests

The **Lagrange multiplier**, or **LM**, test is the third of the three classical tests. The name suggests that it is based on the vector of Lagrange multipliers from a constrained maximization problem. That can indeed be the case. In practice, however, LM tests are very rarely computed in this way. Instead, they are usually based on the gradient vector, or score vector, of the unrestricted loglikelihood function, evaluated at the restricted estimates. LM tests are very often computed by means of artificial regressions. In fact, as we will see, some of the GNR-based tests that we encountered in Sections 6.7 and 7.7 are essentially Lagrange multiplier tests.

For simplicity, we begin our discussion of LM tests by considering the case in which the restrictions to be tested are zero restrictions, that is, restrictions according to which some of the model parameters are zero. In such cases, the r restrictions can be written as $\boldsymbol{\theta}_2 = \mathbf{0}$, where the parameter vector $\boldsymbol{\theta}$ is partitioned as $\boldsymbol{\theta} = [\boldsymbol{\theta}_1 \,\vdots\, \boldsymbol{\theta}_2]$, possibly after some reordering of the elements. The vector $\tilde{\boldsymbol{\theta}}$ of restricted estimates can then be expressed as $\tilde{\boldsymbol{\theta}} = [\tilde{\boldsymbol{\theta}}_1 \,\vdots\, \mathbf{0}]$. The vector $\tilde{\boldsymbol{\theta}}_1$ maximizes the restricted loglikelihood function $\ell(\boldsymbol{\theta}_1, \mathbf{0})$, and so it satisfies the restricted likelihood equations

$$\boldsymbol{g}_1(\tilde{\boldsymbol{\theta}}_1, \mathbf{0}) = \mathbf{0}, \tag{10.63}$$

where $\boldsymbol{g}_1(\cdot)$ is the vector whose components are the $k - r$ partial derivatives of $\ell(\cdot)$ with respect to the elements of $\boldsymbol{\theta}_1$.

The formula (10.38), which gives the asymptotic form of an MLE, can be applied to the estimator $\tilde{\boldsymbol{\theta}}$ when $\boldsymbol{\theta}_2 = \mathbf{0}$. If we partition the true parameter vector $\boldsymbol{\theta}_0$ as $[\boldsymbol{\theta}_1^0 \,\vdots\, \mathbf{0}]$, we find that

$$n^{1/2}(\tilde{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^0) \overset{a}{=} \big(\mathfrak{I}_{11}(\boldsymbol{\theta}_0)\big)^{-1} n^{-1/2} \boldsymbol{g}_1(\boldsymbol{\theta}_0), \tag{10.64}$$

where $\mathfrak{I}_{11}(\cdot)$ is the $(k-r)\times(k-r)$ top left block of the asymptotic information matrix $\mathfrak{I}(\cdot)$ of the full unrestricted model. This block is, of course, just the asymptotic information matrix for the restricted model.