should not be much larger than $\mathrm{SSR}(\hat{\boldsymbol{\beta}})$, or, equivalently, $F/(n-k)$ should be a small quantity. Thus this approximation should generally be a good one. In fact, under the null hypothesis, the LR statistic (10.58) is asymptotically equal to r times the F statistic. Whether or not the null is true, the LR statistic is a deterministic, strictly increasing, function of the F statistic. As we will see later, this fact has important consequences if the statistics are bootstrapped. Without bootstrapping, it makes little sense to use an LR test rather than an F test in the context of the classical normal linear model, because the latter, but not the former, is exact in finite samples.

Wald Tests

Unlike LR tests, Wald tests depend only on the estimates of the unrestricted model. There is no real difference between Wald tests in models estimated by maximum likelihood and those in models estimated by other methods; see Sections 6.7 and 8.5. As with the LR test, we wish to test the r restrictions $\boldsymbol{r}(\boldsymbol{\theta}) = \mathbf{0}$. The Wald test statistic is just a quadratic form in the vector $\boldsymbol{r}(\hat{\boldsymbol{\theta}})$ and the inverse of a matrix that estimates its covariance matrix.

By using the delta method (Section 5.6), we find that

$$\widehat{\mathrm{Var}}\big(\boldsymbol{r}(\hat{\boldsymbol{\theta}})\big) = \boldsymbol{R}(\boldsymbol{\theta}_0)\widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}})\boldsymbol{R}^{\top}(\boldsymbol{\theta}_0), \tag{10.59}$$

where $\boldsymbol{R}(\boldsymbol{\theta})$ is an $r \times k$ matrix with typical element $\partial r_i(\boldsymbol{\theta})/\partial\theta_j$, and $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}})$ is any one of the estimators of $\mathrm{Var}(\hat{\boldsymbol{\theta}})$ that we looked at in the last section. Replacing the unknown $\boldsymbol{\theta}_0$ by $\hat{\boldsymbol{\theta}}$ in (10.59), we find that the Wald statistic is

$$\mathrm{W} = \boldsymbol{r}^{\top}(\hat{\boldsymbol{\theta}})\big(\boldsymbol{R}(\hat{\boldsymbol{\theta}})\widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}})\boldsymbol{R}^{\top}(\hat{\boldsymbol{\theta}})\big)^{-1}\boldsymbol{r}(\hat{\boldsymbol{\theta}}). \tag{10.60}$$

This is a quadratic form in the r–vector $\boldsymbol{r}(\hat{\boldsymbol{\theta}})$, which is asymptotically multivariate normal, and the inverse of an estimate of its covariance matrix. It is easy to see, using the first part of Theorem 4.1, that (10.60) is asymptotically distributed as $\chi^2(r)$ under the null hypothesis. As readers are asked to show in Exercise 10.13, the Wald statistic (6.71) is just a special case of the one defined in (10.60). In the case of linear regression models subject to linear restrictions on the parameters, the Wald statistic is, like the LR statistic, a deterministic, strictly increasing, function of the F statistic if the information matrix estimator (10.43) of the covariance matrix of the parameters is used to construct the Wald statistic.

Wald tests are very widely used, in part because the square of every t statistic is really a Wald statistic. Nevertheless, they should be used with caution. Although Wald tests do not necessarily have poor finite-sample properties, and they do not necessarily perform less well in finite samples than the other classical tests, there is a good deal of evidence that they quite often do so. One reason for this is that Wald statistics are not invariant to reformulations