The result that the statistic $Q(\tilde{\boldsymbol{\beta}}_{\text{FGMM}}, \boldsymbol{y}) - Q(\hat{\boldsymbol{\beta}}_{\text{FGMM}}, \boldsymbol{y})$ is asymptotically distributed as $\chi^2(k_2)$ depends on two critical features of the construction of the statistic. The first is that the same matrix of instruments $\boldsymbol{W}$ is used for estimating both the restricted and unrestricted models. This was also required in Section 8.5, when we discussed testing restrictions on linear regression models estimated by generalized IV. The second essential feature is that the same weighting matrix $(\boldsymbol{W}^\top\hat{\boldsymbol{\Omega}}\boldsymbol{W})^{-1}$ is used when estimating both models. If, as is usually the case, this matrix has to be estimated, it is important that the *same* estimate be used in both criterion functions. If different instruments or different weighting matrices are used for the two models, (9.52) is no longer in general asymptotically distributed as $\chi^2(k_2)$.

One interesting consequence of the form of (9.52) is that we do not always need to bother estimating the unrestricted model. The test statistic (9.52) must always be less than the constrained minimum $Q(\tilde{\boldsymbol{\beta}}_{\text{FGMM}}, \boldsymbol{y})$. Therefore, if $Q(\tilde{\boldsymbol{\beta}}_{\text{FGMM}}, \boldsymbol{y})$ is less than the critical value for the $\chi^2(k_2)$ distribution at our chosen significance level, we can be sure that the actual test statistic is even smaller and would not lead us to reject the null.

The result that tests of restrictions may be based on the difference between the constrained and unconstrained minima of the GMM criterion function holds only for efficient GMM estimation. It is not true for nonoptimal criterion functions like (9.12), which do not use an estimate of the inverse of the covariance matrix of the sample moments as a weighting matrix. When the GMM estimates minimize a nonoptimal criterion function, the easiest way to test restrictions is probably to use a Wald test; see Sections 6.7 and 8.5. However, we do not recommend performing inference on the basis of nonoptimal GMM estimation.

9.5 GMM Estimators for Nonlinear Models

The principles underlying GMM estimation of nonlinear models are the same as those we have developed for GMM estimation of linear regression models. For every result that we have discussed in the previous three sections, there is an analogous result for nonlinear models. In order to develop these results, we will take a somewhat more general and abstract approach than we have done up to this point. This approach, which is based on the theory of **estimating functions**, was originally developed by Godambe (1960) and Durbin (1960); see also Godambe and Thompson (1978).

The method of estimating functions employs the concept of an **elementary zero function**. Such a function plays the same role as a residual in the estimation of a regression model. It depends on observed variables, at least one of which must be endogenous, and on a k–vector of parameters, $\boldsymbol{\theta}$. As with a residual, the expectation of an elementary zero function must vanish if it is evaluated at the true value of $\boldsymbol{\theta}$, but not in general otherwise.