to be numerically identical if the same estimate of the information matrix is used to calculate them. One form, originally proposed by Rao (1948), is called the **score form of the LM test**, or simply the **score test**, and is calculated using the gradient or score vector of the unrestricted model evaluated at the restricted estimates. The other form, which gives the test its name, was proposed by Aitchison and Silvey (1958, 1960) and Silvey (1959). This latter form is calculated using the vector of Lagrange multipliers which emerge if one maximizes the likelihood function subject to constraints by means of a Lagrangian. Econometricians generally use the LM test in its score form but nevertheless insist on calling it an LM test, perhaps because Lagrange multipliers are so widely used in economics. References on LM tests in econometrics include Breusch and Pagan (1980) and Engle (1982a, 1984). Buse (1982) provides an intuitive discussion of the relationships among the LR, LM, and Wald tests.

One way to maximize $\ell(\boldsymbol{\theta})$ subject to the exact restrictions

$$\boldsymbol{r}(\boldsymbol{\theta}) = \mathbf{0}, \tag{8.71}$$

where $\boldsymbol{r}(\boldsymbol{\theta})$ is an r–vector with $r \leq k$, is simultaneously to maximize the Lagrangian

$$\ell(\boldsymbol{\theta}) - \boldsymbol{r}^{\top}(\boldsymbol{\theta})\boldsymbol{\lambda}$$

with respect to $\boldsymbol{\theta}$ and minimize it with respect to the r–vector of Lagrange multipliers $\boldsymbol{\lambda}$. The first-order conditions that characterize the solution to this problem are

$$\begin{aligned} \boldsymbol{g}(\tilde{\boldsymbol{\theta}}) - \boldsymbol{R}^{\top}(\tilde{\boldsymbol{\theta}})\tilde{\boldsymbol{\lambda}} &= \mathbf{0} \\ \boldsymbol{r}(\tilde{\boldsymbol{\theta}}) &= \mathbf{0}, \end{aligned} \tag{8.72}$$

where $\boldsymbol{R}(\boldsymbol{\theta})$ is a $r \times k$ matrix with typical element $\partial r_i(\boldsymbol{\theta})/\partial\theta_j$.

We are interested in the distribution of $\tilde{\boldsymbol{\lambda}}$ under the null hypothesis, so we will suppose that the DGP satisfies (8.71) with parameter vector $\boldsymbol{\theta}_0$. The value of the vector of Lagrange multipliers $\boldsymbol{\lambda}$ if $\tilde{\boldsymbol{\theta}}$ were equal to $\boldsymbol{\theta}_0$ would be zero. Thus it seems natural to take a first-order Taylor expansion of the first-order conditions (8.72) around the point $(\boldsymbol{\theta}_0, \mathbf{0})$. This yields

$$\begin{aligned} \boldsymbol{g}(\boldsymbol{\theta}_0) + \boldsymbol{H}(\bar{\boldsymbol{\theta}})(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - \boldsymbol{R}^{\top}(\bar{\boldsymbol{\theta}})\tilde{\boldsymbol{\lambda}} &= \mathbf{0} \\ \boldsymbol{R}(\ddot{\boldsymbol{\theta}})(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &= \mathbf{0}, \end{aligned}$$

where $\bar{\boldsymbol{\theta}}$ and $\ddot{\boldsymbol{\theta}}$ denote values of $\boldsymbol{\theta}$ that lie between $\tilde{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. These equations may be rewritten as

$$\begin{bmatrix} -\boldsymbol{H}(\bar{\boldsymbol{\theta}}) & \boldsymbol{R}^{\top}(\bar{\boldsymbol{\theta}}) \\ \boldsymbol{R}(\ddot{\boldsymbol{\theta}}) & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \tilde{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{g}(\boldsymbol{\theta}_0) \\ \mathbf{0} \end{bmatrix}. \tag{8.73}$$

If we multiply $\boldsymbol{H}(\bar{\boldsymbol{\theta}})$ by n^{-1}, $\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ by $n^{1/2}$, $\boldsymbol{g}(\boldsymbol{\theta}_0)$ by $n^{-1/2}$, and $\tilde{\boldsymbol{\lambda}}$ by $n^{-1/2}$, we do not change the equality in (8.73), and we render all quantities that