Solution to Exercise 10.13

$\star$**10.13** Consider the linear regression model

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}). \tag{10.106}$$

Derive the Wald statistic for the hypothesis that $\boldsymbol{\beta}_2 = \mathbf{0}$, as a function of the data, from the general formula (10.60). Show that it would be numerically identical to the Wald statistic (6.71) if the same estimate of σ^2 were used.

Show that, if the estimate of σ^2 is either the OLS or the ML estimator based on the unrestricted model (10.106), the Wald statistic is a deterministic, strictly increasing, function of the conventional F statistic. Give the explicit form of this deterministic function. Why can one reasonably expect that this result holds for tests of arbitrary linear restrictions on the parameters, and not only for zero restrictions of the type considered in this exercise?

Expression (10.60) for the Wald statistic is

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

In this case, $\boldsymbol{r}(\hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{\beta}}_2$, and $\boldsymbol{R}(\hat{\boldsymbol{\theta}})$ is the $k_2 \times k$ matrix $[\mathbf{O}\;\;\mathbf{I}]$, where $\mathbf{O}$ is $k_2 \times k_1$ and $\mathbf{I}$ is $k_2 \times k_2$. Therefore, if we interpret $\boldsymbol{\theta}$ as $\boldsymbol{\beta}$, (10.60) becomes

$$\hat{\boldsymbol{\beta}}_2^\top\big([\mathbf{O}\;\;\mathbf{I}]\,\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}})\,[\mathbf{O}\;\;\mathbf{I}]^\top\big)^{-1}\hat{\boldsymbol{\beta}}_2 = \hat{\boldsymbol{\beta}}_2^\top\big(\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_2)\big)^{-1}\hat{\boldsymbol{\beta}}_2. \tag{S10.26}$$

Using the FWL Theorem, it is straightforward to show that

$$\hat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{y},$$

and the ML estimate of its covariance matrix is

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_2) = \hat{\sigma}^2(\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}.$$

Substituting these into the right-hand side of equation (S10.26) for the Wald statistic yields

$$\begin{aligned}\mathrm{W} &= \frac{1}{\hat{\sigma}^2}\boldsymbol{y}^\top\boldsymbol{M}_1\boldsymbol{X}_2(\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}(\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)(\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{y}\\ &= \frac{1}{\hat{\sigma}^2}\boldsymbol{y}^\top\boldsymbol{M}_1\boldsymbol{X}_2(\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{y}. \qquad \text{(S10.27)}\end{aligned}$$

The only difference between this expression and expression (6.71), which was derived in the context of nonlinear least squares, is that the latter uses s^2 instead of $\hat{\sigma}^2$ to estimate σ^2.

The classical F test for $\boldsymbol{\beta}_2 = \mathbf{0}$ in (10.106) is

$$\frac{(\text{RSSR} - \text{USSR})/k_2}{\text{USSR}/(n-k)} = \frac{n-k}{k_2} \times \frac{\boldsymbol{y}^\top \boldsymbol{M}_1 \boldsymbol{X}_2 (\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{y}}{\boldsymbol{y}^\top \boldsymbol{M}_{\boldsymbol{X}} \boldsymbol{y}}. \quad \text{(S10.28)}$$

Since $\hat{\sigma}^2 = \text{USSR}/n$, the Wald statistic (S10.27) is equal to

$$\frac{\boldsymbol{y}^\top \boldsymbol{M}_1 \boldsymbol{X}_2 (\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{y}}{\boldsymbol{y}^\top \boldsymbol{M}_{\boldsymbol{X}} \boldsymbol{y}/n}. \quad \text{(S10.29)}$$

Therefore, from (S10.28) and (S10.29), we see that

$$\text{W} = \frac{nk_2}{n-k} F.$$

The Wald statistic is indeed a deterministic, strictly increasing function of the conventional F statistic, as was to be shown. Note that, if the Wald statistic used s^2 instead of $\hat{\sigma}^2$, the relationship between W and F would be even simpler: $\text{W} = k_2 F$.

Although this result has only been proved for the case of zero restrictions, it is clearly true for arbitrary linear restrictions, because we can always reparametrize a linear regression model so that arbitrary linear restrictions become zero restrictions. This was proved in Exercise 4.8.