Solution to Exercise 15.3

$\star$**15.3** Suppose the dependent variable $\boldsymbol{y}$ is generated by the DGP

$$\boldsymbol{y} = \boldsymbol{X\beta}_0 + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathrm{N}(\boldsymbol{0}, \sigma_0^2\mathbf{I}),$$

where the $n \times k$ matrix $\boldsymbol{X}$ is independent of $\boldsymbol{u}$. Let $\boldsymbol{z}$ be a vector that is not necessarily independent of $\boldsymbol{u}$, but is independent of $\boldsymbol{M_X u}$. Show that the t statistic on $\boldsymbol{z}$ in the linear regression $\boldsymbol{y} = \boldsymbol{X\beta} + c\boldsymbol{z} + \boldsymbol{u}$ follows the Student's t distribution with $n-k-1$ degrees of freedom.

The t statistic can be written as

$$\left(\frac{\boldsymbol{y}^\top \boldsymbol{M}_{\boldsymbol{X},\boldsymbol{z}}\,\boldsymbol{y}}{n-k-1}\right)^{-1/2} \frac{\boldsymbol{z}^\top \boldsymbol{M_X y}}{(\boldsymbol{z}^\top \boldsymbol{M_X z})^{1/2}}.$$

This is just the rightmost expression in equation (4.25) adapted to the test regression in which we are interested. In accordance with our usual notation, $\boldsymbol{M_X}$ denotes the matrix that projects orthogonally onto $\mathbb{S}^\perp(\boldsymbol{X})$, and $\boldsymbol{M}_{\boldsymbol{X},\boldsymbol{z}}$ denotes the matrix that projects orthogonally onto $\mathbb{S}^\perp(\boldsymbol{X}, \boldsymbol{z})$

For the given DGP, $\boldsymbol{M_X y} = \boldsymbol{M_X X\beta}_0 + \boldsymbol{M_X u} = \boldsymbol{M_X u}$, and the t statistic becomes

$$\left(\frac{\boldsymbol{\varepsilon}^\top \boldsymbol{M}_{\boldsymbol{X},\boldsymbol{z}}\,\boldsymbol{\varepsilon}}{n-k-1}\right)^{-1/2} \frac{\boldsymbol{z}^\top \boldsymbol{M_X \varepsilon}}{(\boldsymbol{z}^\top \boldsymbol{M_X z})^{1/2}}, \tag{S15.04}$$

where $\boldsymbol{\varepsilon} \equiv \boldsymbol{u}/\sigma_0$. Notice that expression (S15.04) depends on $\boldsymbol{\varepsilon}$ only through the vector $\boldsymbol{M_X\varepsilon}$. This is the case because $\boldsymbol{M}_{\boldsymbol{X},\boldsymbol{z}}\boldsymbol{\varepsilon} = \boldsymbol{M}_{\boldsymbol{X},\boldsymbol{z}}\boldsymbol{M_X\varepsilon}$, which in turn is true by the result of Exercise 2.15.

We can now reason conditionally on $\boldsymbol{z}$. In order to compute the distribution of the statistic (S15.04) conditional on $\boldsymbol{z}$, we may proceed as though $\boldsymbol{z}$ were deterministic, since $\boldsymbol{z}$ is independent of $\boldsymbol{M_X\varepsilon}$ by hypothesis. But in Section 4.4, in the argument that follows equation (4.25), we showed that, when the regressors are exogenous, a statistic of the form (4.25) follows the $t(n-k)$ distribution, where k is the total number of regressors. Here, in the present notation, this means that, conditionally on $\boldsymbol{z}$, the statistic (S15.04) follows the $t(n-k-1)$ distribution. This distribution does not depend on $\boldsymbol{z}$, and so it is also the unconditional distribution, as we wished to show.