Euclidean length of $\boldsymbol{M_Z x}$ must be smaller (or at least, no larger) than the Euclidean length of $\boldsymbol{x}$; recall (2.27). Thus $\boldsymbol{x}^\top \boldsymbol{M_Z x} \leq \boldsymbol{x}^\top \boldsymbol{x}$, which implies that

$$\sigma_0^2(\boldsymbol{x}^\top \boldsymbol{M_Z x})^{-1} \geq \sigma_0^2(\boldsymbol{x}^\top \boldsymbol{x})^{-1}. \tag{3.57}$$

The inequality in (3.57) almost always holds strictly. The only exception is the special case in which $\boldsymbol{x}$ lies in $\mathcal{S}^\perp(\boldsymbol{Z})$, which implies that the regression of $\boldsymbol{x}$ on $\boldsymbol{Z}$ has no explanatory power at all.

In general, we wish to show that $\mathrm{Var}(\tilde{\boldsymbol{\beta}}) - \mathrm{Var}(\hat{\boldsymbol{\beta}})$ is a positive semidefinite matrix. As we saw in Section 3.5, this is equivalent to showing that the matrix $\mathrm{Var}(\hat{\boldsymbol{\beta}})^{-1} - \mathrm{Var}(\tilde{\boldsymbol{\beta}})^{-1}$ is positive semidefinite. A little algebra shows that

$$\begin{aligned} \boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{X}^\top \boldsymbol{M_Z X} &= \boldsymbol{X}^\top (\mathbf{I} - \boldsymbol{M_Z})\boldsymbol{X} \\ &= \boldsymbol{X}^\top \boldsymbol{P_Z X} \\ &= (\boldsymbol{P_Z X})^\top \boldsymbol{P_Z X}. \end{aligned} \tag{3.58}$$

Since $\boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{X}^\top \boldsymbol{M_Z X}$ can be written as the transpose of a matrix times itself, it must be positive semidefinite. Dividing by σ_0^2 gives the desired result.

We have established that the OLS estimator of $\boldsymbol{\beta}$ in the overspecified regression model (3.51) is at most as efficient as the OLS estimator in the restricted model (3.55), provided the restrictions are true. Therefore, adding additional variables that do not really belong in a model normally leads to less accurate estimates. Only in certain very special cases is there no loss of efficiency. In such cases, the covariance matrices of $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}$ must be the same, which implies that the matrix difference computed in (3.58) must be zero.

The last expression in (3.58) must be a zero matrix whenever $\boldsymbol{P_Z X} = \mathbf{O}$. This condition holds whenever the two sets of regressors $\boldsymbol{X}$ and $\boldsymbol{Z}$ are mutually orthogonal, so that $\boldsymbol{Z}^\top \boldsymbol{X} = \mathbf{O}$. In this special case, $\tilde{\boldsymbol{\beta}}$ is just as efficient as $\hat{\boldsymbol{\beta}}$. In general, however, including regressors that do not belong in a model increases the variance of the estimates of the coefficients on the regressors that do belong, and the increase can be very great in many cases. As can be seen from the left-hand side of (3.57), the variance of the estimated coefficient $\tilde{\beta}$ associated with any regressor $\boldsymbol{x}$ is proportional to the inverse of the SSR from a regression of $\boldsymbol{x}$ on all the other regressors. The more other regressors there are, whether they truly belong in the model or not, the smaller is this SSR.

Underspecification

The opposite of overspecification is underspecification, in which we omit some variables that actually do appear in the DGP. To avoid any new notation, let us suppose that the model we estimate is (3.55), which yields the estimator $\hat{\boldsymbol{\beta}}$, but that the DGP is really

$$\boldsymbol{y} = \boldsymbol{X\beta}_0 + \boldsymbol{Z\gamma}_0 + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathrm{IID}(\mathbf{0}, \sigma_0^2 \mathbf{I}). \tag{3.59}$$