Solution to Exercise 5.11

⋆5.11 Consider a regression model with just two explanatory variables, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, both of which are centered:

$$\boldsymbol{y} = \beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \boldsymbol{u}. \tag{5.13}$$

Let $\hat\rho$ denote the **sample correlation** of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Since both regressors are centered, the sample correlation is

$$\hat\rho \equiv \frac{\sum_{t=1}^n x_{t1}x_{t2}}{\Big((\sum_{t=1}^n x_{t1}^2)(\sum_{t=1}^n x_{t2}^2)\Big)^{1/2}},$$

where x_{t1} and x_{t2} are typical elements of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, respectively. This can be interpreted as the correlation of the joint EDF of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

Show that, under the assumptions of the classical normal linear model, the correlation between the OLS estimates $\hat\beta_1$ and $\hat\beta_2$ is equal to $-\hat\rho$. Which, if any, of the assumptions of this model can be relaxed without changing this result?

The sample correlation $\hat\rho$ can be written in matrix notation as

$$\hat\rho = \frac{\boldsymbol{x}_1^{\top}\boldsymbol{x}_2}{(\boldsymbol{x}_1^{\top}\boldsymbol{x}_1)^{1/2}(\boldsymbol{x}_2^{\top}\boldsymbol{x}_2)^{1/2}}.$$

The correlation between $\hat\beta_1$ and $\hat\beta_2$, conditional on $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, is

$$\frac{\operatorname{Cov}(\hat\beta_1, \hat\beta_2)}{\big(\operatorname{Var}(\hat\beta_1)\operatorname{Var}(\hat\beta_2)\big)^{1/2}}. \tag{S5.14}$$

Using the FWL Theorem, it is not difficult to show that

$$\operatorname{Var}(\hat\beta_1) = \sigma^2(\boldsymbol{x}_1^{\top}\boldsymbol{M}_2\boldsymbol{x}_1)^{-1} \quad \text{and} \quad \operatorname{Var}(\hat\beta_1) = \sigma^2(\boldsymbol{x}_2^{\top}\boldsymbol{M}_1\boldsymbol{x}_2)^{-1},$$

where, as usual, $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ are orthogonal projection matrices. We can rewrite $\boldsymbol{x}_1^{\top}\boldsymbol{M}_2\boldsymbol{x}_1$ as

$$\begin{aligned}
&\boldsymbol{x}_1^{\top}\big(\mathbf{I} - \boldsymbol{x}_2^{\top}(\boldsymbol{x}_2^{\top}\boldsymbol{x}_2)^{-1}\boldsymbol{x}_2^{\top}\big)\boldsymbol{x}_1\\
&\quad = \boldsymbol{x}_1^{\top}\boldsymbol{x}_1 - \boldsymbol{x}_1^{\top}\boldsymbol{x}_2(\boldsymbol{x}_2^{\top}\boldsymbol{x}_2)^{-1}\boldsymbol{x}_2^{\top}\boldsymbol{x}_1\\
&\quad = \left(1 - \frac{(\boldsymbol{x}_1^{\top}\boldsymbol{x}_2)^2}{(\boldsymbol{x}_1^{\top}\boldsymbol{x}_1)(\boldsymbol{x}_2^{\top}\boldsymbol{x}_2)}\right)\boldsymbol{x}_1^{\top}\boldsymbol{x}_1\\
&\quad = (1-\hat\rho^2)\boldsymbol{x}_1^{\top}\boldsymbol{x}_1.
\end{aligned}$$

By the same argument, with the subscripts reversed, we find that

$$\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2 = (1-\hat\rho^2)\,\boldsymbol{x}_2^\top \boldsymbol{x}_2.$$

Thus we conclude that

$$\mathrm{Var}(\hat\beta_1) = \frac{\sigma^2}{1-\hat\rho^2}(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1} \quad \text{and} \quad \mathrm{Var}(\hat\beta_2) = \frac{\sigma^2}{1-\hat\rho^2}(\boldsymbol{x}_2^\top \boldsymbol{x}_2)^{-1}. \tag{S5.15}$$

We now turn our attention to the covariance of $\hat\beta_1$ and $\hat\beta_2$. From standard results, we know that

$$\begin{aligned}
\hat\beta_1 - \beta_1 &= (\boldsymbol{x}_1^\top \boldsymbol{M}_2 \boldsymbol{x}_1)^{-1} \boldsymbol{x}_1^\top \boldsymbol{M}_2 \boldsymbol{u} \quad \text{and} \\
\hat\beta_2 - \beta_2 &= (\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2)^{-1} \boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{u},
\end{aligned}$$

where $\boldsymbol{u}$ is the vector of error terms. Therefore,

$$\begin{aligned}
\mathrm{Cov}(\hat\beta_1, \hat\beta_2) &= \mathrm{E}\big((\hat\beta_1 - \beta_1)(\hat\beta_2 - \beta_2)\big) \\
&= \mathrm{E}\big((\boldsymbol{x}_1^\top \boldsymbol{M}_2 \boldsymbol{x}_1)^{-1} \boldsymbol{x}_1^\top \boldsymbol{M}_2 \boldsymbol{u}\boldsymbol{u}^\top \boldsymbol{M}_1 \boldsymbol{x}_2 (\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2)^{-1}\big) \\
&= \sigma^2 (\boldsymbol{x}_1^\top \boldsymbol{M}_2 \boldsymbol{x}_1)^{-1} \boldsymbol{x}_1^\top \boldsymbol{M}_2 \boldsymbol{M}_1 \boldsymbol{x}_2 (\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2)^{-1}.
\end{aligned}$$

We have already seen that $\boldsymbol{x}_1^\top \boldsymbol{M}_2 \boldsymbol{x}_1 = (1-\hat\rho^2)\,\boldsymbol{x}_1^\top \boldsymbol{x}_1$ and that $\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2 = (1-\hat\rho^2)\,\boldsymbol{x}_2^\top \boldsymbol{x}_2$. Now observe that

$$\begin{aligned}
\boldsymbol{x}_1^\top \boldsymbol{M}_2 \boldsymbol{M}_1 \boldsymbol{x}_2 &= \boldsymbol{x}_1^\top \big(\mathbf{I} - \boldsymbol{x}_2 (\boldsymbol{x}_2^\top \boldsymbol{x}_2)^{-1} \boldsymbol{x}_2^\top\big)\big(\mathbf{I} - \boldsymbol{x}_1 (\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1} \boldsymbol{x}_1^\top\big) \boldsymbol{x}_2 \\
&= \boldsymbol{x}_1^\top \boldsymbol{x}_2 (\boldsymbol{x}_2^\top \boldsymbol{x}_2)^{-1} \boldsymbol{x}_2^\top \boldsymbol{x}_1 (\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1} \boldsymbol{x}_1^\top \boldsymbol{x}_2 - \boldsymbol{x}_1^\top \boldsymbol{x}_2 \\
&= \left(\frac{(\boldsymbol{x}_1^\top \boldsymbol{x}_2)^2}{\boldsymbol{x}_1^\top \boldsymbol{x}_1\, \boldsymbol{x}_2^\top \boldsymbol{x}_2}\right) \boldsymbol{x}_1^\top \boldsymbol{x}_2 - \boldsymbol{x}_1^\top \boldsymbol{x}_2 \\
&= \hat\rho^2\, \boldsymbol{x}_1^\top \boldsymbol{x}_2 - \boldsymbol{x}_1^\top \boldsymbol{x}_2 = (\hat\rho^2 - 1)\, \boldsymbol{x}_1^\top \boldsymbol{x}_2.
\end{aligned}$$

Thus we can write

$$\mathrm{Cov}(\hat\beta_1, \hat\beta_2) = \frac{\sigma^2(\hat\rho^2-1)\,\boldsymbol{x}_1^\top \boldsymbol{x}_2}{(1-\hat\rho^2)^2 \boldsymbol{x}_1^\top \boldsymbol{x}_1\, \boldsymbol{x}_2^\top \boldsymbol{x}_2} = \frac{\sigma^2\, \boldsymbol{x}_1^\top \boldsymbol{x}_2}{(\hat\rho^2-1)\,\boldsymbol{x}_1^\top \boldsymbol{x}_1\, \boldsymbol{x}_2^\top \boldsymbol{x}_2}. \tag{S5.16}$$

Substituting (S5.15) and (S5.16) into (S5.14), we find that the correlation between $\hat\beta_1$ and $\hat\beta_2$ is

$$\frac{(1-\hat\rho^2)\,\boldsymbol{x}_1^\top \boldsymbol{x}_2}{(\hat\rho^2-1)(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{1/2}(\boldsymbol{x}_2^\top \boldsymbol{x}_2)^{1/2}} = -\hat\rho.$$

This completes the proof.

The assumption that the error terms are normally distributed can evidently be relaxed without changing this result, since we never made any use of this assumption. However, the assumption that $\mathrm{E}(\boldsymbol{u}\boldsymbol{u}^\top) = \sigma^2\mathbf{I}$, which we used to obtain $\mathrm{Var}(\hat\beta_1)$, $\mathrm{Var}(\hat\beta_2)$, and $\mathrm{Cov}(\hat\beta_1, \hat\beta_2)$, is evidently essential.