Homework 2

Problem 1. Let $\boldsymbol{X}$ be the matrix collecting all the n observations on the k regressors:

$$\boldsymbol{X} = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,k} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n,1} & X_{n,k} & \cdots & X_{n,k} \end{bmatrix}_{n\times k}.$$

Assume $n > k$ and $\boldsymbol{X}$ is of full rank. Let $\boldsymbol{A}$ be a $k \times k$ *singular* matrix. Show that the columns of $\boldsymbol{XA}$ are linearly dependent and $\mathcal{S}(\boldsymbol{XA}) \subset \mathcal{S}(\boldsymbol{X})$, where

$$\mathcal{S}(\boldsymbol{X}) = \left\{\boldsymbol{z} \in \mathbb{R}^n : \boldsymbol{z} = \boldsymbol{Xb},\, \boldsymbol{b} = (b_1, b_2, \ldots, b_k)' \in \mathbb{R}^k\right\}.$$

Solution. Since $\boldsymbol{A}$ is a $k \times k$ singular matrix, there is at least one k-vector $\boldsymbol{b}$ such that $\boldsymbol{Ab} = \boldsymbol{0}$, and the columns of $\boldsymbol{A}$ must be linearly dependent: let $\boldsymbol{a}_j$ denotes the j-th column of $\boldsymbol{A}$; we have $\boldsymbol{0} = \boldsymbol{Ab} = \begin{bmatrix} \boldsymbol{a}_1 & \ldots & \boldsymbol{a}_k \end{bmatrix}\boldsymbol{b} = b_1\boldsymbol{a}_1 + \ldots + b_k\boldsymbol{a}_k$. Next, the j-th column of $\boldsymbol{XA}$ is given by $\boldsymbol{Xa}_j$, and $\boldsymbol{XAb} = b_1\boldsymbol{Xa}_1 + \ldots + b_k\boldsymbol{Xa}_k$. On the other hand, $\boldsymbol{XAb} = 0$ since $\boldsymbol{Ab} = \boldsymbol{0}$. Therefore, there is a k-vector $\boldsymbol{b}$ such that:

$$b_1\boldsymbol{Xa}_1 + \ldots + b_k\boldsymbol{Xa}_k = \boldsymbol{0}.$$

It follows that the columns of $\boldsymbol{XA}$ are linearly dependent.

To show the second claim, suppose that $\boldsymbol{y} \in \mathcal{S}(\boldsymbol{XA})$. Then there is $\boldsymbol{b} \in \mathbb{R}^k$ such that $\boldsymbol{y} = \boldsymbol{XAb}$. Define $\boldsymbol{c} = \boldsymbol{Ab}$, and note that it is a k-vector. Hence, $\boldsymbol{y} = \boldsymbol{Xc}$, where $\boldsymbol{c} \in \mathbb{R}^k$, and therefore, $\boldsymbol{y} \in \mathcal{S}(\boldsymbol{X})$ by the definition of $\mathcal{S}(\boldsymbol{X})$. We have shown that any $\boldsymbol{y} \in \mathcal{S}(\boldsymbol{XA})$ is also in $\mathcal{S}(\boldsymbol{X})$. Hence, $\mathcal{S}(\boldsymbol{XA}) \subset \mathcal{S}(\boldsymbol{X})$.

Problem 2. Partition the matrix of regressors $\boldsymbol{X}$ as follows:

$$\boldsymbol{X} = [\boldsymbol{X}_1 \ \ \boldsymbol{X}_2].$$

Denote $\boldsymbol{P}_1 = \boldsymbol{X}_1\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'$ and $\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'$. $\boldsymbol{M}_1$ and $\boldsymbol{M}_{\boldsymbol{X}}$ are defined analogously: $\boldsymbol{M}_1 = \boldsymbol{I}_n - \boldsymbol{P}_1$ and $\boldsymbol{M}_{\boldsymbol{X}} = \boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}}$. Prove:

$$\boldsymbol{P}_1\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{P}_1 = \boldsymbol{P}_1 \tag{1}$$

and

$$\boldsymbol{M}_1\boldsymbol{M}_{\boldsymbol{X}} = \boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{M}_1 = \boldsymbol{M}_{\boldsymbol{X}}. \tag{2}$$

Solution. Since $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{X}_1 = \boldsymbol{X}_1$,

$$\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{P}_1 = \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{X}_1\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1' = \boldsymbol{X}_1\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1' = \boldsymbol{P}_1.$$

Transpose:

$$\boldsymbol{P}_1 = \boldsymbol{P}_1' = (\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{P}_1)' = \boldsymbol{P}_1'\boldsymbol{P}_{\boldsymbol{X}}' = \boldsymbol{P}_1\boldsymbol{P}_{\boldsymbol{X}}. \tag{3}$$

Then,

$$\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{M}_1 = (\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})(\boldsymbol{I}_n - \boldsymbol{P}_1) = \boldsymbol{I}_n - \boldsymbol{P}_1 - \boldsymbol{P}_{\boldsymbol{X}} + \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{P}_1 = \boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{M}_{\boldsymbol{X}}.$$

$\boldsymbol{M}_1\boldsymbol{M}_{\boldsymbol{X}} = \boldsymbol{M}_{\boldsymbol{X}}$ follows from steps similar to (3).

Problem 3. Partition the matrix of regressors $\boldsymbol{X}$ as follows:

$$\boldsymbol{X} = [\boldsymbol{X}_1 \;\; \boldsymbol{X}_2].$$

Consider the regression model

$$\boldsymbol{Y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{e}.$$

Let $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ denote the LS estimates from running the regression:

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}}_1 \\ \widehat{\boldsymbol{\beta}}_2 \end{pmatrix} = \underset{(\boldsymbol{b}_1,\boldsymbol{b}_2)\in\mathbb{R}^{k_1}\times\mathbb{R}^{k_2}}{\operatorname{argmin}} (\boldsymbol{Y} - \boldsymbol{X}_1\boldsymbol{b}_1 - \boldsymbol{X}_2\boldsymbol{b}_2)' (\boldsymbol{Y} - \boldsymbol{X}_1\boldsymbol{b}_1 - \boldsymbol{X}_2\boldsymbol{b}_2).$$

Denote $\widehat{\boldsymbol{e}} = \boldsymbol{Y} - \boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2$.

Consider the following regressions, all to be estimated by LS:

(i) Regress $\boldsymbol{Y}$ on $\boldsymbol{X}_2$. Let $\widetilde{\boldsymbol{\beta}}_2$ denote the LS estimates.

(ii) Regress $\boldsymbol{P}_1\boldsymbol{Y}$ on $\boldsymbol{X}_2$. Let $\widetilde{\boldsymbol{\beta}}_2$ denote the LS estimates.

(iii) Regress $\boldsymbol{P}_1\boldsymbol{Y}$ on $\boldsymbol{P}_1\boldsymbol{X}_2$. Let $\widetilde{\boldsymbol{\beta}}_2$ denote the LS estimates.

(iv) Regress $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{Y}$ on $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. Let $\begin{pmatrix} \widetilde{\boldsymbol{\beta}}_1 \\ \widetilde{\boldsymbol{\beta}}_2 \end{pmatrix}$ denote the LS estimates.

(v) Regress $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{Y}$ on $\boldsymbol{X}_2$. Let $\widetilde{\boldsymbol{\beta}}_2$ denote the LS estimates.

(vi) Regress $\boldsymbol{M}_1\boldsymbol{Y}$ on $\boldsymbol{X}_2$. Let $\widetilde{\boldsymbol{\beta}}_2$ denote the LS estimates.

(vii) Regress $\boldsymbol{M}_1\boldsymbol{Y}$ on $\boldsymbol{M}_1\boldsymbol{X}_2$. Let $\widetilde{\boldsymbol{\beta}}_2$ denote the LS estimates.

(viii) Regress $\boldsymbol{M}_1\boldsymbol{Y}$ on $\boldsymbol{X}_1$ and $\boldsymbol{M}_1\boldsymbol{X}_2$. Let $\begin{pmatrix} \widetilde{\boldsymbol{\beta}}_1 \\ \widetilde{\boldsymbol{\beta}}_2 \end{pmatrix}$ denote the LS estimates.

(ix) Regress $\boldsymbol{M}_1\boldsymbol{Y}$ on $\boldsymbol{M}_1\boldsymbol{X}_1$ and $\boldsymbol{M}_1\boldsymbol{X}_2$. Let $\begin{pmatrix} \widetilde{\boldsymbol{\beta}}_1 \\ \widetilde{\boldsymbol{\beta}}_2 \end{pmatrix}$ denote the LS estimates.

(x) Regress $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{Y}$ on $\boldsymbol{M}_1\boldsymbol{X}_2$. Let $\widetilde{\boldsymbol{\beta}}_2$ denote the LS estimates.

For which of the above regressions will the estimates be the same as $\widehat{\boldsymbol{\beta}}_2$? For which will the residuals be the same as $\widehat{\boldsymbol{e}}$?

Solution. Note that by FWL theorem,$\widehat{\boldsymbol{\beta}}_2 = \left(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2\right)^{-1} \boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{Y}$ and the residuals are

$$\widehat{\boldsymbol{e}} = \boldsymbol{M}_1\boldsymbol{Y} - \boldsymbol{M}_1\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 = \boldsymbol{M}_1\boldsymbol{Y} - \boldsymbol{M}_1\boldsymbol{X}_2\left(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2\right)^{-1} \boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{Y} = \boldsymbol{M}_1(\boldsymbol{I}_n - \boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{M}_1)\boldsymbol{Y}.$$

(i) $\widetilde{\boldsymbol{\beta}}_2 = \left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1} \boldsymbol{X}_2'\boldsymbol{Y}$. In general, neither the estimates nor the residuals are the same as the original regression, because $\boldsymbol{X}_2$ does not span the same space as $\boldsymbol{X}$. If $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are orthogonal ($\boldsymbol{X}_1'\boldsymbol{X}_2 = \boldsymbol{0}$), then $\widehat{\boldsymbol{\beta}}_2 = \widetilde{\boldsymbol{\beta}}_2$.

(ii) $\widetilde{\boldsymbol{\beta}}_2 = \left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1} \boldsymbol{X}_2'\boldsymbol{P}_1\boldsymbol{Y} \neq \widehat{\boldsymbol{\beta}}_2$, therefore the estimate for β_2 is not the same as the original regression. The residual is probably not the same either, because neither the dependent variable nor the independent variables are exactly the same as the original regression.

(iii) $\widetilde{\boldsymbol{\beta}}_2 = \left(\boldsymbol{X}_2'\boldsymbol{P}_1\boldsymbol{X}_2\right)^{-1} \boldsymbol{X}_2'\boldsymbol{P}_1\boldsymbol{Y} \neq \widehat{\boldsymbol{\beta}}_2$, therefore the estimate for β_2 is not the same as the original regression. The residual is probably not the same either, because neither the dependent variable nor the independent variables are exactly the same as the original regression.

(iv) Since

$$\begin{aligned}
\widetilde{\boldsymbol{\beta}}_2 &= (\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{Y} \\
&= (\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'(\boldsymbol{I}_n - \boldsymbol{P}_1)\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{Y} \\
&= (\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\left(\boldsymbol{X}_2'\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{X}_2'\boldsymbol{P}_1\boldsymbol{P}_{\boldsymbol{X}}\right)\boldsymbol{Y} \\
&= (\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\left(\boldsymbol{X}_2' - \boldsymbol{X}_2'\boldsymbol{P}_1\right)\boldsymbol{Y} \\
&= (\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\left(\boldsymbol{I}_n - \boldsymbol{P}_1\right)\boldsymbol{Y} \\
&= \widehat{\boldsymbol{\beta}}_2,
\end{aligned}$$

the estimates are the same, while the residuals are $\mathbf{0}$ in this case, because the dependent variable is the fitted value of the original regression. (It has been used above that $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{X}_2 = \boldsymbol{X}_2$ and that $\boldsymbol{P}_1\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{P}_1$.)

(v) $\widetilde{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{Y} \neq \widehat{\boldsymbol{\beta}}_2$, therefore the estimate for β_2 is not the same as the original regression. The residual is probably not the same either, because neither the dependent variable nor the independent variables are exactly the same as the original regression.

(vi) $\widetilde{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{Y} \neq \widehat{\boldsymbol{\beta}}_2$, therefore the estimate for β_2 is not the same as the original regression. The residual is probably not the same either, because neither the dependent variable nor the independent variables are exactly the same as the original regression.

(vii) Both the estimate and the residual are the same by the FWL Theorem.

(viii) Both the estimate and the residual are the same by the FWL Theorem.

(ix) Note $\boldsymbol{M}_1\boldsymbol{X}_1 = \mathbf{0}$. So the space spanned by the columns of $\boldsymbol{M}_1\boldsymbol{X}_1$ and $\boldsymbol{M}_1\boldsymbol{X}_2$ is the same as that spanned by only $\boldsymbol{M}_1\boldsymbol{X}_2$. So $\widetilde{\boldsymbol{\beta}}_2$ is the same as the estimate from regressing $\boldsymbol{M}_1\boldsymbol{Y}$ on $\boldsymbol{M}_1\boldsymbol{X}_2$. This is a confusing question, because the data actually violates the "no multicollinearity" assumption. In this case $\widetilde{\boldsymbol{\beta}}_1$ is actually not determined. But it can be seen that the residual is also the same as $\widehat{\boldsymbol{e}}$.

(x) $\widetilde{\boldsymbol{\beta}}_2 = \left(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{Y}$ and it can be shown that

$$\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{Y} = \boldsymbol{X}_2'\left(\boldsymbol{I}_n - \boldsymbol{P}_1\right)\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{Y} = (\boldsymbol{X}_2'\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{X}_2'\boldsymbol{P}_1)\boldsymbol{Y} = (\boldsymbol{X}_2' - \boldsymbol{X}_2'\boldsymbol{P}_1)y = \boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{Y}.$$

Therefore the estimates are the same, while the residuals may not be the same. The residuals from regressing $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{Y}$ on $\boldsymbol{M}_1\boldsymbol{X}_2$ are

$$\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{Y} - \boldsymbol{M}_1\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 = (\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{Y} - \boldsymbol{M}_1\boldsymbol{Y}) + \left(\boldsymbol{M}_1\boldsymbol{Y} - \boldsymbol{M}_1\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2\right) = (\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{Y} - \boldsymbol{M}_1\boldsymbol{Y}) + \widehat{\boldsymbol{e}}.$$

But in general, $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{Y} \neq \boldsymbol{M}_1\boldsymbol{Y}$.

Problem 4. Consider the linear regression

$$\boldsymbol{Y} = \mathbf{1}\beta_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{e}.$$

Use the FWL theorem to show that the LS estimates of β_1 and $\boldsymbol{\beta}_2$ can be written as

$$\begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\boldsymbol{\beta}}_2 \end{pmatrix} = \begin{pmatrix} n & \mathbf{1}'\boldsymbol{X}_2 \\ 0 & \boldsymbol{X}_2'\boldsymbol{M}_{\mathbf{1}}\boldsymbol{X}_2 \end{pmatrix}^{-1}\begin{pmatrix} \mathbf{1}'\boldsymbol{Y} \\ \boldsymbol{X}_2'\boldsymbol{M}_{\mathbf{1}}\boldsymbol{Y} \end{pmatrix}.$$

$\boldsymbol{M}_{\mathbf{1}} = \boldsymbol{I}_n - \mathbf{1}\left(\mathbf{1}'\mathbf{1}\right)^{-1}\mathbf{1}'$.

Solution. Using the FWL theorem, we have:

$$\widehat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{Y},$$

i.e.,

$$(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2)\widehat{\boldsymbol{\beta}}_2 = \boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{Y}. \tag{4}$$

In an OLS regression, we have :

$$\boldsymbol{1}'\widehat{\boldsymbol{e}} = \boldsymbol{1}'(\boldsymbol{Y} - \widehat{\beta}_1\boldsymbol{1} - \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2) = 0,$$

i.e.,

$$\boldsymbol{1}'\boldsymbol{Y} = \hat{\beta}_1\boldsymbol{1}'\boldsymbol{1} + \boldsymbol{1}'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2.$$

Therefore,

$$n\hat{\beta}_1 + \boldsymbol{1}'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 = \boldsymbol{1}'\boldsymbol{Y} \tag{5}$$

Combing equation system (1) and (2), we have the desired results.

Problem 5. Use (1) to show that $\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1$ is symmetric and idempotent. Show further that $\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1 = \boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}$ by showing that for any $\boldsymbol{z} \in \mathcal{S}(\boldsymbol{M}_1\boldsymbol{X}_2)$, $(\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1)\boldsymbol{z} = \boldsymbol{z}$ and for any $\boldsymbol{y} \in \mathcal{S}^{\perp}(\boldsymbol{M}_1\boldsymbol{X}_2)$, $(\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1)\boldsymbol{y} = \boldsymbol{0}$, where

$$\mathcal{S}^{\perp}(\boldsymbol{M}_1\boldsymbol{X}_2) = \{\boldsymbol{z} \in \mathbb{R}^n : \boldsymbol{z}'\boldsymbol{M}_1\boldsymbol{X}_2 = \boldsymbol{0}\}.$$

Solution. We have to show that $(\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1)$ is symmetric and idempotent. a) symmetric: since both $\boldsymbol{P}_{\boldsymbol{X}}$ and $\boldsymbol{P}_1$ are symmetric, $\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1$ is also symmetric. b) idempotent:

$$(\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1)(\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1) = \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{P}_1 - \boldsymbol{P}_1\boldsymbol{P}_{\boldsymbol{X}} + P_1P_1 = \boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1 - \boldsymbol{P}_1 + \boldsymbol{P}_1 = \boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1.$$

Take any $\boldsymbol{z} \in \mathcal{S}(\boldsymbol{M}_1\boldsymbol{X}_2)$, then $\boldsymbol{z}$ can be written as $\boldsymbol{z} = \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\alpha}$ for some vector $\boldsymbol{\alpha}$.

$$(\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1)\boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\alpha} = (-\boldsymbol{M}_{\boldsymbol{X}} + \boldsymbol{M}_1)\boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\alpha} = -\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{X}_2\boldsymbol{\alpha} + \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\alpha} = \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\alpha} = \boldsymbol{z},$$

where we used $\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{M}_1 = \boldsymbol{M}_{\boldsymbol{X}}$ and $\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{X}_2 = \boldsymbol{0}$.

Suppose $\boldsymbol{y}'\boldsymbol{M}_1\boldsymbol{X}_2 = \boldsymbol{0}$. Then,

$$\boldsymbol{y}'\boldsymbol{M}_1\boldsymbol{X} = \boldsymbol{y}'\boldsymbol{M}_1[\boldsymbol{X}_1 \ \ \boldsymbol{X}_2] = [\boldsymbol{y}'\boldsymbol{M}_1\boldsymbol{X}_1 \ \ \boldsymbol{y}'\boldsymbol{M}_1\boldsymbol{X}_2] = \boldsymbol{0},$$

since $\boldsymbol{M}_1\boldsymbol{X}_1 = \boldsymbol{0}$. Transpose to get

$$\boldsymbol{0} = \boldsymbol{X}'\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{X}'(\boldsymbol{I}_n - \boldsymbol{P}_1)\boldsymbol{y} \Longrightarrow \boldsymbol{X}'\boldsymbol{y} = \boldsymbol{X}'\boldsymbol{P}_1\boldsymbol{y}.$$

Then, premultiply by $\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$:

$$\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{P}_1\boldsymbol{y} \Longrightarrow \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} = \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{P}_1\boldsymbol{y} \Longrightarrow \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} = \boldsymbol{P}_1\boldsymbol{y} \Longrightarrow (\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1)\boldsymbol{y} = \boldsymbol{0},$$

where we used (1).

Problem 6. Consider two regressions. (a) Regress $\boldsymbol{Y}$ on $\boldsymbol{X}$:

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{b}\in\mathbb{R}^{k_1}}{\operatorname{argmin}}\,(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b}).$$

(b) Regress $\boldsymbol{Y}$ on $\boldsymbol{X}$ and $\boldsymbol{Z}$:

$$\begin{pmatrix} \widetilde{\boldsymbol{\beta}} \\ \widetilde{\gamma} \end{pmatrix} = \underset{(\boldsymbol{b},\boldsymbol{c})\in\mathbb{R}^{k_1}\times\mathbb{R}^{k_2}}{\operatorname{argmin}}\,(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b} - \boldsymbol{Z}\boldsymbol{c})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b} - \boldsymbol{Z}\boldsymbol{c}).$$

Show that in the case of mutually orthogonal regressors, with $\boldsymbol{X}'\boldsymbol{Z} = \boldsymbol{0}$, $\widehat{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}$.

Solution. By FWL theorem,

$$\begin{aligned}\widetilde{\boldsymbol{\beta}} &= \left(\boldsymbol{X}'\left(\boldsymbol{I}_n - \boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\right)\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\left(\boldsymbol{I}_n - \boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\right)\boldsymbol{Y}\right) \\ &= \left(\boldsymbol{X}'\boldsymbol{X} - \boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{Y} - \boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Y}\right) \\ &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y} = \widehat{\boldsymbol{\beta}}.\end{aligned}$$

Problem 7. In this question, use the hints to show "R^2 increases by adding more regressors". Suppose we have n observations on regressors $(Z_1, ..., Z_k)$ and $(W_1, ..., W_m)$ and dependent variable Y. Let $\boldsymbol{Z}$ be the $n \times k$ matrix collecting the observations on $(Z_1, ..., Z_k)$ and let $\boldsymbol{W}$ be the $n \times m$ matrix collecting the observations on $(W_1, ..., W_m)$. Let $\boldsymbol{X} = [\boldsymbol{Z} \;\; \boldsymbol{W}]$. Assume that $\boldsymbol{Z}$ contains a column of ones so that $R^2 = 1 - RSS/TSS$ in both regressions.

Let

$$\boldsymbol{P_X} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' \text{ projection matrix corresponding to the full regression,}$$
$$\boldsymbol{P_Z} = \boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}' \text{ projection matrix corresponding to the regression without } \boldsymbol{W}.$$

Define also

$$\begin{aligned}\boldsymbol{M_X} &= \boldsymbol{I}_n - \boldsymbol{P_X}, \\ \boldsymbol{M_Z} &= \boldsymbol{I}_n - \boldsymbol{P_Z}.\end{aligned}$$

Define

$$\begin{aligned}\widehat{\boldsymbol{e}}_{\boldsymbol{X}} &= \boldsymbol{M_X}\boldsymbol{Y}, \\ \widehat{\boldsymbol{e}}_{\boldsymbol{Z}} &= \boldsymbol{M_Z}\boldsymbol{Y}.\end{aligned}$$

Show: $\widehat{\boldsymbol{e}}_{\boldsymbol{X}}'\widehat{\boldsymbol{e}}_{\boldsymbol{Z}} = \widehat{\boldsymbol{e}}_{\boldsymbol{X}}'\widehat{\boldsymbol{e}}_{\boldsymbol{X}}$ and therefore

$$0 \leq \left(\widehat{\boldsymbol{e}}_{\boldsymbol{X}} - \widehat{\boldsymbol{e}}_{\boldsymbol{Z}}\right)'\left(\widehat{\boldsymbol{e}}_{\boldsymbol{X}} - \widehat{\boldsymbol{e}}_{\boldsymbol{Z}}\right) = \widehat{\boldsymbol{e}}_{\boldsymbol{X}}'\widehat{\boldsymbol{e}}_{\boldsymbol{X}} - \widehat{\boldsymbol{e}}_{\boldsymbol{Z}}'\widehat{\boldsymbol{e}}_{\boldsymbol{Z}}.$$

Hint: Use (1) and (2). How can you argue that now we conclude that "R^2 increases by adding more regressors"?

Solution. Note that since $\boldsymbol{Z}$ is a part of $\boldsymbol{X}$,

$$\boldsymbol{P_X}\boldsymbol{Z} = \boldsymbol{Z},$$

and

$$\begin{aligned}\boldsymbol{P_X}\boldsymbol{P_Z} &= \boldsymbol{P_X}\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}' \\ &= \boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}' \\ &= \boldsymbol{P_Z}.\end{aligned}$$

Consequently,

$$\begin{aligned}\boldsymbol{M_X}\boldsymbol{M_Z} &= \left(\boldsymbol{I}_n - \boldsymbol{P_X}\right)\left(\boldsymbol{I}_n - \boldsymbol{P_Z}\right) \\ &= \boldsymbol{I}_n - \boldsymbol{P_X} - \boldsymbol{P_Z} + \boldsymbol{P_X}\boldsymbol{P_Z} \\ &= \boldsymbol{I}_n - \boldsymbol{P_X} - \boldsymbol{P_Z} + \boldsymbol{P_Z} \\ &= \boldsymbol{M_X}.\end{aligned}$$

Assume that $\boldsymbol{Z}$ contains a column of ones, so both short and long regressions have intercepts. Define

$$\begin{aligned}\widehat{\boldsymbol{e}}_{\boldsymbol{X}} &= \boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{Y},\\ \widehat{\boldsymbol{e}}_{\boldsymbol{Z}} &= \boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{Y}.\end{aligned}$$

Write:

$$\begin{aligned}0 &\le \left(\widehat{\boldsymbol{e}}_{\boldsymbol{X}} - \widehat{\boldsymbol{e}}_{\boldsymbol{Z}}\right)'\left(\widehat{\boldsymbol{e}}_{\boldsymbol{X}} - \widehat{\boldsymbol{e}}_{\boldsymbol{Z}}\right)\\ &= \widehat{\boldsymbol{e}}_{\boldsymbol{X}}'\widehat{\boldsymbol{e}}_{\boldsymbol{X}} + \widehat{\boldsymbol{e}}_{\boldsymbol{Z}}'\widehat{\boldsymbol{e}}_{\boldsymbol{Z}} - 2\widehat{\boldsymbol{e}}_{\boldsymbol{X}}'\widehat{\boldsymbol{e}}_{\boldsymbol{Z}}.\end{aligned}$$

Next,

$$\begin{aligned}\widehat{\boldsymbol{e}}_{\boldsymbol{X}}'\widehat{\boldsymbol{e}}_{\boldsymbol{Z}} &= \boldsymbol{Y}'\boldsymbol{M}_X\boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{Y}\\ &= \boldsymbol{Y}'\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{Y}\\ &= \widehat{\boldsymbol{e}}_{\boldsymbol{X}}'\widehat{\boldsymbol{e}}_{\boldsymbol{X}}.\end{aligned}$$

Hence,

$$\widehat{\boldsymbol{e}}_{\boldsymbol{Z}}'\widehat{\boldsymbol{e}}_{\boldsymbol{Z}} \ge \widehat{\boldsymbol{e}}_{\boldsymbol{X}}'\widehat{\boldsymbol{e}}_{\boldsymbol{X}}.$$

Note that $\widehat{\boldsymbol{e}}_{\boldsymbol{Z}}'\widehat{\boldsymbol{e}}_{\boldsymbol{Z}}$ is the RSS of the short regression and $\widehat{\boldsymbol{e}}_{\boldsymbol{X}}'\widehat{\boldsymbol{e}}_{\boldsymbol{X}}$ is the RSS of the long regression and the two regressions have the same TSS. Since $R^2 = 1 - RSS/TSS$, comparing R^2 is equivalent to comparing RSS.

Problem 8. Consider the following linear regression model:

$$\begin{aligned}Y_i &= X_{1,i}\beta_1 + X_{2,i}\beta_2 + e_i\\ \mathbb{E}\begin{pmatrix} X_{1,i}\\ X_{2,i}\end{pmatrix} e_i &= \mathbf{0}\end{aligned}$$

with n observations. Consider the constraint

$$\beta_1 = 2\beta_2.$$

Find the expression of the constrained LS estimator under this constraint:

$$\min_{b_1,b_2}\sum_{i=1}^{n}\left(Y_i - X_{1,i}b_1 - X_{2,i}b_2\right)^2 \text{ subject to } b_1 = 2b_2$$

Solution. The constrained LS estimator:

$$\left(\widehat{\beta}_1, \widehat{\beta}_2\right) = \underset{b_1,b_2}{\operatorname{argmin}}\sum_{i=1}^{n}\left(Y_i - X_{1,i}b_1 - X_{2,i}b_2\right)^2 \text{ subject to } b_1 = 2b_2.$$

Note

$$\widehat{\beta}_2 = \underset{b_2}{\operatorname{argmin}}\sum_{i=1}^{n}\left(Y_i - 2X_{1,i}b_2 - X_{2,i}b_2\right)^2$$

and $\widehat{\beta}_1 = 2\widehat{\beta}_2$. Solving the first order condition,

$$\widehat{\beta}_2 = \frac{\sum_{i=1}^{n}\left(2X_{1,i} + X_{2,i}\right)Y_i}{\sum_{i=1}^{n}\left(2X_{1,i} + X_{2,i}\right)^2}.$$

Problem 9. Let $\boldsymbol{X}$ be an $n \times k$ matrix ($n > k$) of full column rank. Partition $\boldsymbol{X}$ as $\boldsymbol{X} = \begin{bmatrix}\boldsymbol{X}_1 & \boldsymbol{X}_2\end{bmatrix}$, where $\boldsymbol{X}_1$ is $n \times k_1$ and $\boldsymbol{X}_2$ is $n \times k_2$, $k_1 + k_2 = k$.

(i) Show that $\boldsymbol{X}_2$ has full column rank and therefore $(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}$ exists.

(ii) Define $\boldsymbol{M}_2 = \boldsymbol{I}_n - \boldsymbol{X}_2 \left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1} \boldsymbol{X}_2'$ and $\widetilde{\boldsymbol{X}}_1 = \boldsymbol{M}_2\boldsymbol{X}_1$. Show that $\widetilde{\boldsymbol{X}}_1$ has full column rank and therefore $\left(\widetilde{\boldsymbol{X}}_1'\widetilde{\boldsymbol{X}}_1\right)^{-1} = \left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)^{-1}$ exists.

Solution. (i)Proof by contradiction: Suppose $\boldsymbol{X}_2$ does not have full rank, then there exist a vector $\boldsymbol{\lambda}$ such that $\boldsymbol{X}_2\boldsymbol{\lambda} = \mathbf{0}$. Then we have:

$$\begin{pmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\lambda} \end{pmatrix} = 0, \text{ where} \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\lambda} \end{pmatrix} \neq \mathbf{0}.$$

This is contradict to the given condition that $\boldsymbol{X}$ has full column rank. Then $\boldsymbol{X}_2'\boldsymbol{X}_2$ is a $k_2 \times k_2$ matrix, and rank($\boldsymbol{X}_2'\boldsymbol{X}_2$)= rank $(\boldsymbol{X}_2) = k_2$, which is full rank, then $(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}$ exist.
(ii). Proof by contradiction: Suppose $\widetilde{\boldsymbol{X}}_1 = \boldsymbol{M}_2\boldsymbol{X}_1$ does not have full column rank, then there exist a nonzero vector $\boldsymbol{\beta}$ such that $\widetilde{\boldsymbol{X}}_1\boldsymbol{\beta} = \mathbf{0}$, i.e.

$$\begin{aligned} & \boldsymbol{M}_2\boldsymbol{X}_1\boldsymbol{\beta} = (\boldsymbol{I}_n - \boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2')\boldsymbol{X}_1\boldsymbol{\beta} = \mathbf{0} \\ \iff & \boldsymbol{X}_1\boldsymbol{\beta} - \boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{X}_1\boldsymbol{\beta} = \mathbf{0} \\ \iff & \begin{pmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ -(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{X}_1\boldsymbol{\beta} \end{pmatrix} = \mathbf{0} \end{aligned}$$

where $\begin{pmatrix} \boldsymbol{\beta} \\ -(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{X}_1\boldsymbol{\beta} \end{pmatrix} \neq \mathbf{0}$. This contradict to that $\boldsymbol{X}$ has full rank. Therefore, $\widetilde{\boldsymbol{X}}_1$ has full column rank, and $(\widetilde{\boldsymbol{X}}_1'\widetilde{\boldsymbol{X}}_1)^{-1}$ exist.

A direct proof: We want to show that the $n \times k_1$ matrix $\boldsymbol{M}_2\boldsymbol{X}_1$ has full column rank, i.e. rank$(\boldsymbol{M}_2\boldsymbol{X}_1) = k_1$. First, rank$(\boldsymbol{M}_2\boldsymbol{X}_1) \leq \min\{\text{rank}(\boldsymbol{M}_2), \text{rank}(\boldsymbol{X}_1)\}$. It can be shown that rank$(\boldsymbol{M}_2) = n - k_2$ and rank$(\boldsymbol{X}_1) = k_1$. Since $k_1 + k_2 \leq n$, $k_1 \leq n - k_2$, rank$(\boldsymbol{M}_2\boldsymbol{X}_1) \leq$ rank$(\boldsymbol{X}_1) = k_1$. Second, observe that rank$(\boldsymbol{M}_2\boldsymbol{X}_1) =$ rank$(\boldsymbol{M}_2\boldsymbol{X})$ since $\boldsymbol{M}_2\boldsymbol{X}_2 = \mathbf{0}$. Then, by Sylvester-inequality, rank$(\boldsymbol{M}_2\boldsymbol{X}_1) =$ rank$(\boldsymbol{M}_2\boldsymbol{X}) \geq$ rank$(\boldsymbol{M}_2)$ + rank$(\boldsymbol{X}) - n = n - k_2 + k - n = k_1$. Combining the previous two results, rank$(\boldsymbol{M}_2\boldsymbol{X}) = k_1$.

Problem 10. Consider a partitioned linear regression model $\boldsymbol{Y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{e}$, and assume that the assumptions of the classical normal linear regression hold. Let $\widehat{\boldsymbol{\beta}}_1$ be the OLS estimator of $\boldsymbol{\beta}_1$. Show that $\text{Var}(\widehat{\boldsymbol{\beta}}_1|\boldsymbol{X}_1, \boldsymbol{X}_2) = \sigma^2(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}$, where $\boldsymbol{M}_2 = \boldsymbol{I}_n - \boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'$.

Solution.

$$\begin{aligned} \text{Var}(\widehat{\boldsymbol{\beta}}_1|\boldsymbol{X}_1, \boldsymbol{X}_2) =& \text{Var}((\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{Y})|\boldsymbol{X}_1, \boldsymbol{X}_2) \\ =& \text{Var}((\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}(\boldsymbol{X}_1'\boldsymbol{M}_2(\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{e}))|\boldsymbol{X}_1, \boldsymbol{X}_2) \\ =& \text{Var}((\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)\boldsymbol{\beta}_1 + (\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}(X_1'M_2X_2)\boldsymbol{\beta}_2 + (\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}(X_1'M_2e)|\boldsymbol{X}_1, \boldsymbol{X}_2) \\ =& \text{Var}(\boldsymbol{\beta}_1 + \mathbf{0} + (\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{e})|\boldsymbol{X}_1, \boldsymbol{X}_2) \\ =& (\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{M}_2\text{Var}(\boldsymbol{e}|\boldsymbol{X}_1, \boldsymbol{X}_2)M_2X_1(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1} \\ =& (\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{M}_2\sigma^2\boldsymbol{I}_n\boldsymbol{M}_2\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1} \\ =& \sigma^2(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{M}_2\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1} \\ =& \sigma^2(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1} \end{aligned}$$

where $(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}(X_1'M_2X_2)\boldsymbol{\beta}_2 = \mathbf{0}$ since $\boldsymbol{M}_2\boldsymbol{X}_2 = \mathbf{0}$.

Problem 11. Let $\boldsymbol{X}$ be the matrix collecting all the n observations on the k regressors. Let $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{B}$, where $\boldsymbol{B}$ is a $k \times k$ non-singular matrix. Let $\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{e}}\right)$ denote the LS estimates and residuals from regression of $\boldsymbol{Y}$ on $\boldsymbol{X}$. Similarly, let $\left(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{e}}\right)$ denote these from regression of $\boldsymbol{Y}$ on $\boldsymbol{Z}$. Find the relationship between $\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{e}}\right)$ and $\left(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{e}}\right)$.

Solution.

$$
\begin{aligned}
\widetilde{\boldsymbol{\beta}} &= \left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Y} \\
&= \left(\boldsymbol{B}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{B}\right)^{-1}\boldsymbol{B}'\boldsymbol{X}'\boldsymbol{Y} \\
&= \boldsymbol{B}^{-1}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{B}'\right)^{-1}\boldsymbol{B}'\boldsymbol{X}'\boldsymbol{Y} \\
&= \boldsymbol{B}^{-1}\widehat{\boldsymbol{\beta}}.
\end{aligned}
$$

$$
\begin{aligned}
\widetilde{\boldsymbol{e}} &= \left(\boldsymbol{I}_n - \boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\right)\boldsymbol{Y} \\
&= \left(\boldsymbol{I}_n - \boldsymbol{X}\boldsymbol{B}\left(\boldsymbol{B}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{B}\right)^{-1}\boldsymbol{B}'\boldsymbol{X}\right)\boldsymbol{Y} \\
&= \left(\boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}\right)\boldsymbol{Y} \\
&= \widehat{\boldsymbol{e}}.
\end{aligned}
$$

We used $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.

Problem 12. Suppose that assumptions of the Classical Linear Regression model hold, i.e.

$$
\begin{aligned}
\boldsymbol{Y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e},\ \boldsymbol{\beta} \in \mathbb{R}^k \\
\mathbb{E}(\boldsymbol{e}|\boldsymbol{X}) &= 0, \\
\operatorname{rank}(\boldsymbol{X}) &= k,
\end{aligned}
$$

however,

$$\mathbb{E}(\boldsymbol{e}\boldsymbol{e}'|\boldsymbol{X}) = \boldsymbol{\Omega},$$

where $\boldsymbol{\Omega}$ is an $n \times n$, positive definite and symmetric matrix, but different from $\sigma^2\boldsymbol{I}_n$.

(i) Derive the conditional variance (given $\boldsymbol{X}$) of the LS estimator $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$.

(ii) Derive the conditional variance (given $\boldsymbol{X}$) of the Generalized LS estimator $\widetilde{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{Y}$.

(iii) Without relying on the Gauss-Markov Theorem, show that

$$\operatorname{Var}(\widehat{\boldsymbol{\beta}} \mid \boldsymbol{X}) - \operatorname{Var}(\widetilde{\boldsymbol{\beta}} \mid \boldsymbol{X}) \geq 0$$

(in the positive semidefinite sense). Hint: Show

$$\left(\operatorname{Var}(\widetilde{\boldsymbol{\beta}} \mid \boldsymbol{X})\right)^{-1} - \left(\operatorname{Var}(\widehat{\boldsymbol{\beta}} \mid \boldsymbol{X})\right)^{-1} \geq 0$$

by showing that the expression on the left-hand side depends on a symmetric and idempotent matrix of the form $\boldsymbol{I}_n - \boldsymbol{H}(\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'$ for some $n \times k$ matrix $\boldsymbol{H}$ of rank k.

Solution.

(i) Recall, $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{e}$ where $\boldsymbol{\beta}$ is a constant and does not vary. Therefore,

$$
\begin{aligned}
\operatorname{Var}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) &= \operatorname{Var}((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{e}|\boldsymbol{X}) \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\operatorname{Var}(\boldsymbol{e}|\boldsymbol{X})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbb{E}(\boldsymbol{e}\boldsymbol{e}'|\boldsymbol{X})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}
\end{aligned}
$$

(ii) As above, $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{e}$. Therefore,

$$\begin{aligned}\mathrm{Var}(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X}) &= \mathrm{Var}((\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{e}|\boldsymbol{X}) \\ &= (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\mathrm{Var}(\boldsymbol{e}|\boldsymbol{X})\boldsymbol{\Omega}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1} \\ &= (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Omega}\boldsymbol{\Omega}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1} \\ &= (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1} \\ &= (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}.\end{aligned}$$

(iii) Following the hint we first show,

$$\begin{aligned}&(\mathrm{Var}(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X}))^{-1} - (\mathrm{Var}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}))^{-1} \\ =&\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X} - (\boldsymbol{X}'\boldsymbol{X})\left(\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X}\right)^{-1}(\boldsymbol{X}'\boldsymbol{X}) \\ =&\boldsymbol{X}'\boldsymbol{\Omega}^{-1/2}(\boldsymbol{I}_n - \boldsymbol{\Omega}^{1/2}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{1/2})\boldsymbol{\Omega}^{-1/2}\boldsymbol{X} \\ =&\boldsymbol{X}'\boldsymbol{\Omega}^{-1/2}(\boldsymbol{I}_n - \boldsymbol{D}(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}')\boldsymbol{\Omega}^{-1/2}\boldsymbol{X}\end{aligned}$$

where $\boldsymbol{D} = \boldsymbol{\Omega}^{1/2}\boldsymbol{X}$. Notice then that $\boldsymbol{I}_n - \boldsymbol{D}(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'$ is a symmetric idempotent matrix. Since any symmetric idempotent matrix is positive definite we have that,

$$(\mathrm{Var}(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X}))^{-1} - (\mathrm{Var}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}))^{-1} \geq 0 \Rightarrow \mathrm{Var}(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X}) - \mathrm{Var}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) \leq 0.$$

Problem 13. Consider the GLS estimator $\widetilde{\boldsymbol{\beta}}$ defined in the previous question.

(i) Show that $\widetilde{\boldsymbol{\beta}}$ satisfies $\widetilde{\boldsymbol{e}}'\boldsymbol{\Omega}^{-1}\boldsymbol{X} = \boldsymbol{0}$, where $\widetilde{\boldsymbol{e}} = \boldsymbol{Y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}$.

(ii) Using the result in (i), show that the generalized squared distance function $S(\boldsymbol{b}) = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b})'\boldsymbol{\Omega}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b})$ can be written as

$$S(\boldsymbol{b}) = \widetilde{\boldsymbol{e}}'\boldsymbol{\Omega}^{-1}\widetilde{\boldsymbol{e}} + (\widetilde{\boldsymbol{\beta}} - \boldsymbol{b})'\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{b}).$$

(iii) Using the result in (ii), show that $\widetilde{\boldsymbol{\beta}}$ minimizes $S(\boldsymbol{b})$.

Solution.

(i)

$$\begin{aligned}\widetilde{\boldsymbol{e}}'\boldsymbol{\Omega}^{-1}\boldsymbol{X} &= (\boldsymbol{Y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}})'\boldsymbol{\Omega}^{-1}\boldsymbol{X} \\ &= \boldsymbol{Y}'\boldsymbol{\Omega}^{-1}\boldsymbol{X} - \widetilde{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X} \\ &= \boldsymbol{Y}'\boldsymbol{\Omega}^{-1}\boldsymbol{X} - \left[\boldsymbol{Y}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\right]\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X} \\ &= \boldsymbol{0}.\end{aligned}$$

(ii) By adding and subtracting $\boldsymbol{X}\widetilde{\boldsymbol{\beta}}$, we have

$$\begin{aligned}Q(\boldsymbol{b}) &= (\boldsymbol{Y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}} + \boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{b})'\boldsymbol{\Omega}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}} + \boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{b}) \\ &= (\widetilde{\boldsymbol{e}} + \boldsymbol{X}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{b}))'\boldsymbol{\Omega}^{-1}(\widetilde{\boldsymbol{e}} + \boldsymbol{X}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{b})) \\ &= \widetilde{\boldsymbol{e}}'\boldsymbol{\Omega}^{-1}\widetilde{\boldsymbol{e}} + (\widetilde{\boldsymbol{\beta}} - \boldsymbol{b})'\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{b}) + 2\widetilde{\boldsymbol{e}}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{b}).\end{aligned}$$

However, $\widetilde{\boldsymbol{e}}'\boldsymbol{\Omega}^{-1}\boldsymbol{X} = \boldsymbol{0}$ due to the result in part (i).

(iii) $\widetilde{\boldsymbol{e}}'\boldsymbol{\Omega}^{-1}\widetilde{\boldsymbol{e}}$ does not depend on $\boldsymbol{b}$, and therefore minimization of $Q(\boldsymbol{b})$ is equivalent to minimization of $(\widetilde{\boldsymbol{\beta}} - \boldsymbol{b})'\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{b})$. Since $\boldsymbol{\Omega}$ is positive definite, $\boldsymbol{\Omega}^{-1}$ and $\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}$ are also positive definite as the rank of $\boldsymbol{X}$ is k. Hence,

$$(\widetilde{\boldsymbol{\beta}} - \boldsymbol{b})'\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{b}) \geq 0$$

which holds as an equality for $\boldsymbol{b} = \widetilde{\boldsymbol{\beta}}$.