Homework 2

Problem 1. Let $Y \in \{1,2,3,4,5,6\}$ be the face number showing when a die is rolled. Define X as

$$X = \begin{cases} Y & \text{if } Y \text{ is even}, \\ 0 & \text{if } Y \text{ is odd}. \end{cases}$$

Find the best linear predictor $\mathcal{P}(Y|X)$ and the conditional expectation $\mathbb{E}(Y|X)$. Calculate $\mathbb{E}\left[(Y - \mathcal{P}(Y|X))^2\right]$ and $\mathbb{E}\left[(Y - \mathbb{E}(Y|X))^2\right]$.

Solution. First we derive the joint distribution of (X,Y):

$$\begin{aligned} \frac{1}{6} =& \Pr((X,Y)=(0,1)) \\ =& \Pr((X,Y)=(0,3)) \\ =& \Pr((X,Y)=(0,5)) \\ =& \Pr((X,Y)=(2,2)) \\ =& \Pr((X,Y)=(4,4)) \\ =& \Pr((X,Y)=(6,6)). \end{aligned}$$

It is easy to derive that the best linear predictor is:

$$\mathcal{P}(Y|X) = \mathbb{E}(Y) + \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}(X - \mathbb{E}(X)).$$

We have

$$\begin{aligned} \mathbb{E}(Y) &= \sum_{i=1}^{6} \frac{i}{6} = \frac{7}{2} \\ \mathbb{E}(X) &= \frac{1}{2} \times 0 + \frac{1}{6}(2+4+6) = 2 \\ \operatorname{Var}(X) &= \frac{1}{2} \times (0-2)^2 + \frac{1}{6}\left[(2-2)^2 + (4-2)^2 + (6-2)^2\right] = \frac{16}{3} \\ \mathbb{E}(XY) &= \frac{1}{6}\left(2^2 + 4^2 + 6^2\right) = \frac{28}{3} \\ \operatorname{Cov}(X,Y) &= \frac{28}{3} - 2 \times \frac{28}{3} = \frac{7}{3}. \end{aligned}$$

So

$$\mathcal{P}(Y|X) = \frac{21}{8} + \frac{7}{16}X.$$

The conditional mean is:

$$\mathbb{E}(Y|X) = \begin{cases} X & \text{if } X = 2,4,6 \\ \frac{1}{3}(1+3+5) & \text{if } X = 0. \end{cases}$$

Calculate:

$$\mathbb{E}\left((Y - \mathcal{P}(Y|X))^2\right) = \frac{1}{6}\left[\left(1 - \frac{21}{8}\right)^2 + \left(3 - \frac{21}{8}\right)^2 + \left(5 - \frac{21}{8}\right)^2 + \left(2 - \left(\frac{21}{8} + \frac{7}{16} \times 2\right)\right)^2\right.$$

$$
\begin{aligned}
&+\left(4-\left(\frac{21}{8}+\frac{7}{16}\times 4\right)\right)^2+\left(6-\left(\frac{21}{8}+\frac{7}{16}\times 6\right)\right)^2\Bigg] \\
&=1.896
\end{aligned}
$$

and

$$\mathbb{E}\left[(Y-\mathbb{E}(Y|X))^2\right] = \frac{1}{6}\left[(1-3)^2+(3-3)^2+(5-3)^2\right]=1.333.$$

Problem 2. Suppose that

$$
\begin{aligned}
Y =& \boldsymbol{X}'\boldsymbol{\beta}+e \\
\mathrm{E}(e|\boldsymbol{X}) =& 0 \\
\mathrm{E}\left(e^2|\boldsymbol{X}\right) =& \sigma^2(\boldsymbol{X}).
\end{aligned}
$$

Consider two approximations to the conditional variance $\sigma^2(\boldsymbol{X})$:

$$\boldsymbol{\gamma}_1 \text{ minimizes } \mathrm{E}\left(\sigma^2(\boldsymbol{X})-\boldsymbol{X}'\gamma\right)^2$$

and

$$\boldsymbol{\gamma}_2 \text{ minimizes } \mathrm{E}\left(e^2-\boldsymbol{X}'\gamma\right)^2.$$

Show: $\boldsymbol{\gamma}_1=\boldsymbol{\gamma}_2$.

Solution. By law of iterated expectation,

$$
\begin{aligned}
\boldsymbol{\gamma}_2 &= \left(\mathbb{E}(\boldsymbol{X}\boldsymbol{X}')\right)^{-1}\mathbb{E}\left(\boldsymbol{X}e^2\right) \\
&= \left(\mathbb{E}(\boldsymbol{X}\boldsymbol{X}')\right)^{-1}\mathbb{E}\left(\boldsymbol{X}\mathbb{E}\left(e^2|\boldsymbol{X}\right)\right) \\
&= \left(\mathbb{E}(\boldsymbol{X}\boldsymbol{X}')\right)^{-1}\mathbb{E}\left(\boldsymbol{X}\sigma^2(\boldsymbol{X})\right) \\
&= \boldsymbol{\gamma}_1.
\end{aligned}
$$

Problem 3. The Mean Trimmed Squared Error (MTSE) is defined by

$$T(\boldsymbol{\theta})=\mathrm{E}\left(\left(Y-\boldsymbol{X}'\boldsymbol{\theta}\right)^2\tau(\boldsymbol{X})\right),$$

where $\tau(\boldsymbol{X})$ is a known, scalar-valued, non-negative, bounded, function.

1. Give an explicit formula for the value of $\boldsymbol{\theta}$ which minimizes $T(\boldsymbol{\theta})$.
2. Define $e=Y-\boldsymbol{X}'\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is the minimizer defined above. Show: $\mathrm{E}(\boldsymbol{X}\tau(\boldsymbol{X})e)=0$.
3. Under what condition (other than $\tau(\boldsymbol{X})=1$) will this minimizer equal the Best Linear Predictor?

Solution. Part (1). By expanding the square

$$
\begin{aligned}
T(\boldsymbol{\theta}) &= \mathbb{E}\left(\left(Y-\boldsymbol{X}'\boldsymbol{\theta}\right)^2\tau(\boldsymbol{X})\right) \\
&= \mathbb{E}\left(Y^2\tau(\boldsymbol{X})\right)-2\mathbb{E}\left(Y\boldsymbol{X}'\tau(\boldsymbol{X})\right)\boldsymbol{\theta}+\boldsymbol{\theta}'\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\tau(\boldsymbol{X})\right)\boldsymbol{\theta}.
\end{aligned}
$$

Differentiate:

$$\frac{\partial}{\partial\boldsymbol{\theta}}T(\boldsymbol{\theta})=-2\mathbb{E}(\boldsymbol{X}Y\tau(\boldsymbol{X}))+2\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\tau(\boldsymbol{X})\right)\boldsymbol{\theta}.$$

Setting it equal to zero and solving for $\boldsymbol{\theta}$:

$$\boldsymbol{\theta}=\left(\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\tau(\boldsymbol{X})\right)\right)^{-1}\mathbb{E}(\boldsymbol{X}Y\tau(\boldsymbol{X})).$$

Part (2). Since $e = Y - \boldsymbol{X}'\boldsymbol{\theta}$,

$$\begin{aligned}
\mathbb{E}\left(\boldsymbol{X}e\tau\left(\boldsymbol{X}\right)\right) &= \mathbb{E}\left(\boldsymbol{X}Y\tau\left(\boldsymbol{X}\right)\right) - \mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\tau\left(\boldsymbol{X}\right)\right)\boldsymbol{\theta} \\
&= \mathbb{E}\left(\boldsymbol{X}Y\tau\left(\boldsymbol{X}\right)\right) - \mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\tau\left(\boldsymbol{X}\right)\right)\left(\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\tau\left(\boldsymbol{X}\right)\right)\right)^{-1}\mathbb{E}\left(\boldsymbol{X}Y\tau\left(\boldsymbol{X}\right)\right) \\
&= 0.
\end{aligned}$$

Part (3). If the conditional mean is linear: $\mathbb{E}\left(Y|\boldsymbol{X}\right) = \boldsymbol{X}'\boldsymbol{\beta}$ then by the law of iterated expectation,

$$\begin{aligned}
\boldsymbol{\theta} &= \left(\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\tau\left(\boldsymbol{X}\right)\right)\right)^{-1}\mathbb{E}\left(\boldsymbol{X}Y\tau\left(\boldsymbol{X}\right)\right) \\
&= \left(\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\tau\left(\boldsymbol{X}\right)\right)\right)^{-1}\mathbb{E}\mathbb{E}\left(\boldsymbol{X}Y\tau\left(\boldsymbol{X}\right)|\boldsymbol{X}\right) \\
&= \left(\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\tau\left(\boldsymbol{X}\right)\right)\right)^{-1}\mathbb{E}\boldsymbol{X}\tau\left(\boldsymbol{X}\right)\mathbb{E}\left(Y|\boldsymbol{X}\right) \\
&= \left(\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\tau\left(\boldsymbol{X}\right)\right)\right)^{-1}\mathbb{E}\left(\boldsymbol{X}\tau\left(\boldsymbol{X}\right)\boldsymbol{X}'\right)\boldsymbol{\beta} \\
&= \boldsymbol{\beta}.
\end{aligned}$$

Problem 4. The conditional PDF of Y given X is

$$f_{Y|X}\left(y|x\right) = \frac{y+x}{1/2+x},$$

for $0 < y < 1$. Find $\mathrm{E}\left[Y|X=x\right]$.

Solution.

$$\mathrm{E}\left[Y|X=x\right] = \int_0^1 \frac{y\left(y+x\right)}{1/2+x}\mathrm{d}y = \frac{2+3x}{3+6x}.$$

Problem 5. For any given two random variables X and Y, we define

$$\mathrm{Var}\left[Y \mid X\right] = \mathrm{E}\left[\left(Y - \mathrm{E}\left[Y \mid X\right]\right)^2 \mid X\right].$$

Suppose that $\mathrm{E}\left[Y \mid X\right] = 1/4$ and $\mathrm{E}\left[Y^2 \mid X\right] = 1/8$. Then show that for any function g, $\mathrm{Var}\left[Y \mid g\left(X\right)\right] = 1/16$. Use the following facts: for any function g, $\mathrm{E}\left[\mathrm{E}\left[Y \mid g\left(X\right)\right] \mid X\right] = \mathrm{E}\left[Y \mid g\left(X\right)\right]$ and $\mathrm{E}\left[\mathrm{E}\left[Y \mid X\right] \mid g\left(X\right)\right] = \mathrm{E}\left[Y \mid g\left(X\right)\right]$.

Solution. By using the fact that $\mathrm{E}\left[Y \mid g\left(X\right)\right] = \mathrm{E}\left[\mathrm{E}\left[Y \mid X\right] \mid g\left(X\right)\right]$,

$$\begin{aligned}
\mathrm{Var}\left[Y \mid g\left(X\right)\right] &= \mathrm{E}\left[\left(Y - \mathrm{E}\left[Y \mid g\left(X\right)\right]\right)^2 \mid g\left(X\right)\right] \\
&= \mathrm{E}\left[\left(Y - \mathrm{E}\left[\mathrm{E}\left[Y \mid X\right] \mid g\left(X\right)\right]\right)^2 \mid g\left(X\right)\right] \\
&= \mathrm{E}\left[\left(Y - \frac{1}{4}\right)^2 \mid g\left(X\right)\right] \\
&= \mathrm{E}\left[Y^2 \mid g\left(X\right)\right] - \frac{1}{2}\mathrm{E}\left[Y \mid g\left(X\right)\right] + \mathrm{E}\left[\frac{1}{16} \mid g\left(X\right)\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[Y^2 \mid X\right] \mid g\left(X\right)\right] - \frac{1}{2}\mathrm{E}\left[\mathrm{E}\left[Y \mid X\right] \mid g\left(X\right)\right] + \frac{1}{16} \\
&= \mathrm{E}\left[\frac{1}{8} \mid g\left(X\right)\right] - \frac{1}{2}\mathrm{E}\left[\frac{1}{4} \mid g\left(X\right)\right] + \frac{1}{16} \\
&= \frac{1}{8} - \frac{1}{8} + \frac{1}{16} \\
&= \frac{1}{16}.
\end{aligned}$$

Problem 6. 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 $\boldsymbol{0} = \boldsymbol{Ab} = [\ \boldsymbol{a}_1 \ \ \ldots \ \ \boldsymbol{a}_k\ ]\boldsymbol{b} = b_1\boldsymbol{a}_1 + \ldots + b_k\boldsymbol{a}_k$. 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 7. 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 $\boldsymbol{P_X} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'$. $\boldsymbol{M}_1$ and $\boldsymbol{M_X}$ are defined analogously: $\boldsymbol{M}_1 = \boldsymbol{I}_n - \boldsymbol{P}_1$ and $\boldsymbol{M_X} = \boldsymbol{I}_n - \boldsymbol{P_X}$. Prove:

$$\boldsymbol{P}_1\boldsymbol{P_X} = \boldsymbol{P_X}\boldsymbol{P}_1 = \boldsymbol{P}_1 \tag{1}$$

and

$$\boldsymbol{M}_1\boldsymbol{M_X} = \boldsymbol{M_X}\boldsymbol{M}_1 = \boldsymbol{M_X}. \tag{2}$$

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

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

Transpose:

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

Then,

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

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

Problem 8. Use (1) to show that $\boldsymbol{P_X} - \boldsymbol{P}_1$ is symmetric and idempotent. Show further that $\boldsymbol{P_X} - \boldsymbol{P}_1 = \boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}$ by showing that for any $\boldsymbol{z} \in \mathcal{S}(\boldsymbol{M}_1\boldsymbol{X}_2)$, $(\boldsymbol{P_X} - \boldsymbol{P}_1)\boldsymbol{z} = \boldsymbol{z}$ and for any $\boldsymbol{y} \in \mathcal{S}^{\perp}(\boldsymbol{M}_1\boldsymbol{X}_2)$, $(\boldsymbol{P_X} - \boldsymbol{P}_1)\boldsymbol{y} = \boldsymbol{0}$, 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 $(\boldsymbol{P_X} - \boldsymbol{P}_1)$ is symmetric and idempotent. a) symmetric: since both $\boldsymbol{P_X}$ and $\boldsymbol{P}_1$ are symmetric, $\boldsymbol{P_X} - \boldsymbol{P}_1$ is also symmetric. b) idempotent:

$$(\boldsymbol{P_X} - \boldsymbol{P}_1)(\boldsymbol{P_X} - \boldsymbol{P}_1) = \boldsymbol{P_X}\boldsymbol{P_X} - \boldsymbol{P_X}\boldsymbol{P}_1 - \boldsymbol{P}_1\boldsymbol{P_X} + \boldsymbol{P}_1\boldsymbol{P}_1 = \boldsymbol{P_X} - \boldsymbol{P}_1 - \boldsymbol{P}_1 + \boldsymbol{P}_1 = \boldsymbol{P_X} - \boldsymbol{P}_1.$$

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 $\boldsymbol{\alpha}$.

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

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

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 $\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$:

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

where we used (1).

Problem 9. 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

$$\boldsymbol{M_X} = \boldsymbol{I}_n - \boldsymbol{P_X},$$
$$\boldsymbol{M_Z} = \boldsymbol{I}_n - \boldsymbol{P_Z}.$$

Define

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

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

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

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

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

Consequently,

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

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

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

Write:

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

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,

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

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 10. Let $\boldsymbol{X}$ be an $n \times k$ matrix $(n > k)$ of full column rank. Partition $\boldsymbol{X}$ as $\boldsymbol{X} = \left[\begin{array}{ll}\boldsymbol{X}_1 & \boldsymbol{X}_2\end{array}\right]$, where $\boldsymbol{X}_1$ is $n \times k_1$ and $\boldsymbol{X}_2$ is $n \times k_2$, $k_1 + k_2 = k$.

1. Show that $\boldsymbol{X}_2$ has full column rank and therefore $(\boldsymbol{X}'_2\boldsymbol{X}_2)^{-1}$ exists.
2. 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 $\boldsymbol{X}_2\boldsymbol{\lambda} = \boldsymbol{0}$. Then we have:

$$\left(\begin{array}{ll}\boldsymbol{X}_1 & \boldsymbol{X}_2\end{array}\right)\left(\begin{array}{c}\boldsymbol{0} \\ \boldsymbol{\lambda}\end{array}\right) = 0, \text{ where } \left(\begin{array}{c}\boldsymbol{0} \\ \boldsymbol{\lambda}\end{array}\right) \neq \boldsymbol{0}.$$

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. $\text{rank}(\boldsymbol{M}_2\boldsymbol{X}_1) = k_1$. First, $\text{rank}(\boldsymbol{M}_2\boldsymbol{X}_1) \leq \min\{\text{rank}(\boldsymbol{M}_2), \text{rank}(\boldsymbol{X}_1)\}$. It can be shown that $\text{rank}(\boldsymbol{M}_2) = n - k_2$ and $\text{rank}(\boldsymbol{X}_1) = k_1$. Since $k_1 + k_2 \leq n$, $k_1 \leq n - k_2$, $\text{rank}(\boldsymbol{M}_2\boldsymbol{X}_1) \leq \text{rank}(\boldsymbol{X}_1) = k_1$. Second, observe that $\text{rank}(\boldsymbol{M}_2\boldsymbol{X}_1) = \text{rank}(\boldsymbol{M}_2\boldsymbol{X})$ since $\boldsymbol{M}_2\boldsymbol{X}_2 = \mathbf{0}$. Then, by Sylvester-inequality, $\text{rank}(\boldsymbol{M}_2\boldsymbol{X}_1) = \text{rank}(\boldsymbol{M}_2\boldsymbol{X}) \geq \text{rank}(\boldsymbol{M}_2) + \text{rank}(\boldsymbol{X}) - n = n - k_2 + k - n = k_1$. Combining the previous two results, $\text{rank}(\boldsymbol{M}_2\boldsymbol{X}) = k_1$.