# Homework 2

**Problem 1.** 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?

**Problem 2.** 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

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

**Problem 3.** 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

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

**Problem 4.** 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

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

**Problem 5.** 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

$\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X} \left( \boldsymbol{X}'\boldsymbol{X} \right)^{-1} \boldsymbol{X}'$ projection matrix corresponding to the full regression,

$\boldsymbol{P}_{\boldsymbol{Z}} = \boldsymbol{Z} \left( \boldsymbol{Z}'\boldsymbol{Z} \right)^{-1} \boldsymbol{Z}'$ projection matrix corresponding to the regression without $\boldsymbol{W}$.

Define also

$$\boldsymbol{M}_{\boldsymbol{X}} = \boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}},$$
$$\boldsymbol{M}_{\boldsymbol{Z}} = \boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{Z}}.$$

Define

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

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

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

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

**Problem 6.** Let $\boldsymbol{X}$ be an $n \times k$ matrix ($n > k$) of full column rank. Partition $\boldsymbol{X}$ as $\boldsymbol{X} = [\ \boldsymbol{X}_1 \quad \boldsymbol{X}_2\ ]$, 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.

**Problem 7.** Suppose that assumptions of the Classical Linear Regression model hold, i.e.

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

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 $\sigma^2\boldsymbol{I}_n$.

1. Derive the conditional variance (given $\boldsymbol{X}$) of the LS estimator $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$.
2. Derive the conditional variance (given $\boldsymbol{X}$) of the Generalized LS estimator $\widetilde{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{Y}$.
3. Without relying on the Gauss-Markov Theorem, show that

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

(in the positive semidefinite sense). Hint: Show

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

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 $k$.

**Problem 8.** Consider the GLS estimator $\widetilde{\boldsymbol{\beta}}$ defined in the previous question.

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

2. 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
$$S(\boldsymbol{b}) = \widetilde{\boldsymbol{e}}'\boldsymbol{\Omega}^{-1}\widetilde{\boldsymbol{e}} + (\widetilde{\boldsymbol{\beta}} - \boldsymbol{b})'\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{b}).$$

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

**Problem 9.** Consider the following regression model:

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{e}, \\ \mathbb{E}(\boldsymbol{e}|\boldsymbol{X}_1, \boldsymbol{X}_2) &= 0, \\ \mathbb{E}\left(\boldsymbol{e}\boldsymbol{e}'|\boldsymbol{X}_1, \boldsymbol{X}_2\right) &= \sigma_e^2 \boldsymbol{I}_n. \end{aligned}$$

Let $\widetilde{\boldsymbol{\beta}}_1 = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{Y}$ be the LS estimator for $\boldsymbol{\beta}_1$ which omits $\boldsymbol{X}_2$ from the regression.

1. Find $\mathbb{E}(\tilde{\boldsymbol{\beta}}_1|\boldsymbol{X}_1)$.

2. Define
$$\boldsymbol{V} = \boldsymbol{X}_2\boldsymbol{\beta}_2 - \mathbb{E}\left(\boldsymbol{X}_2\boldsymbol{\beta}_2|\boldsymbol{X}_1\right).$$
Find $\mathbb{E}\left(\boldsymbol{e}\boldsymbol{V}'|\boldsymbol{X}_1\right)$.

3. Find $\mathbb{E}\left(\boldsymbol{e}\boldsymbol{e}'|\boldsymbol{X}_1\right)$.

4. Assume that
$$\mathbb{E}\left(\boldsymbol{V}\boldsymbol{V}'|\boldsymbol{X}_1\right) = \sigma_v^2 I_n,$$
and find $\mathrm{Var}(\tilde{\boldsymbol{\beta}}_1|\boldsymbol{X}_1)$.

5. Let $\hat{\boldsymbol{\beta}}_1 = (\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{Y}$ be the OLS estimator for $\boldsymbol{\beta}_1$ from a regression of $\boldsymbol{Y}$ against $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, where $\boldsymbol{M}_2 = \boldsymbol{I}_n - \boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'$. Compare $\mathrm{Var}(\tilde{\boldsymbol{\beta}}_1|\boldsymbol{X}_1)$ derived in part (iv) with $\mathrm{Var}(\hat{\boldsymbol{\beta}}_1|\boldsymbol{X}_1, \boldsymbol{X}_2)$. Can you say which of the two variances is bigger (in the positive semi-definite sense)? Explain your answer.