Midterm Exam

Problem 1. (10 Points) Let Y and X be two random variables.

(i) Show that $\mathbb{E}(Y|X)$ and $Y - \mathbb{E}(Y|X)$ are uncorrelated. **Hint**: Use law of iterated expectations.

(ii) Show that $\mathrm{Var}(Y) \geq \mathrm{Var}(Y - \mathbb{E}(Y|X))$. **Hint**: Use (i).

Solution. (i) $\mathbb{E}\left(Y - \mathbb{E}(Y|X)\right) = \mathbb{E}Y - \mathbb{E}\mathbb{E}\left(Y|X\right) = 0$. Then,

$$\begin{aligned}
\mathrm{Cov}\left(\mathbb{E}(Y|X), Y - \mathbb{E}(Y|X)\right) &= \mathbb{E}\left(\mathbb{E}(Y|X) \cdot (Y - \mathbb{E}(Y|X))\right) - \left(\mathbb{E}\mathbb{E}\left(Y|X\right)\right)\left(\mathbb{E}\left(Y - \mathbb{E}(Y|X)\right)\right) \\
&= \mathbb{E}\left(\mathbb{E}(Y|X) \cdot (Y - \mathbb{E}(Y|X))\right) \\
&= \mathbb{E}\mathbb{E}\left(\mathbb{E}(Y|X) \cdot (Y - \mathbb{E}(Y|X))\,|X\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(Y|X\right) \cdot \mathbb{E}\left(Y - \mathbb{E}(Y|X)|X\right)\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(Y|X\right)\left(\mathbb{E}\left(Y|X\right) - \mathbb{E}\left(Y|X\right)\right)\right) \\
&= 0.
\end{aligned}$$

(ii)

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathrm{Var}\left(\mathbb{E}(Y|X) + (Y - \mathbb{E}(Y|X))\right) \\
&= \mathrm{Var}\left(\mathbb{E}(Y|X)\right) + \mathrm{Var}\left(Y - \mathbb{E}(Y|X)\right) + 2 \cdot \mathrm{Cov}\left(\mathbb{E}(Y|X), Y - \mathbb{E}(Y|X)\right) \\
&= \mathrm{Var}\left(\mathbb{E}(Y|X)\right) + \mathrm{Var}\left(Y - \mathbb{E}(Y|X)\right) \\
&\geq \mathrm{Var}\left(Y - \mathbb{E}(Y|X)\right).
\end{aligned}$$

Problem 2. (20 Points) Y and X are both (scalar) random variables. The Mean Trimmed Squared Error (MTSE) is defined by

$$T(\theta) = \mathbb{E}\left((Y - \theta X)^2 \tau(X)\right),$$

where $\tau(X)$ is a known, scalar-valued, non-negative, bounded, function. Suppose there is a (i.i.d.) random sample $\{(Y_i, X_i)\}_{i=1}^n$.

(i) Give an explicit formula for the value of θ which minimizes $T(\theta)$.

(ii) Define $e = Y - X\theta$, where θ is the minimizer defined above. Show: $\mathbb{E}\left(X\tau(X)e\right) = 0$.

(iii) Write down the method of moments estimator $\widehat{\theta}$ of θ.

(iv) Show that $\widehat{\theta}$ is consistent for θ.

(v) (**Bonus Question**, 5 Points) Find the asymptotic distribution of $\sqrt{n}\left(\hat{\theta} - \theta\right)$.

Solution.

(i) By expanding the square

$$\begin{aligned}
T(\theta) &= \mathbb{E}\left((Y - \theta X)^2 \tau(X)\right) \\
&= \mathbb{E}\left(Y^2 \tau(X)\right) - 2\mathbb{E}\left(YX\tau(X)\right)\theta + \mathbb{E}\left(X^2\tau(X)\right)\theta^2.
\end{aligned}$$

Differentiate:

$$\frac{\partial}{\partial\theta} T(\theta) = -2\mathbb{E}\left(YX\tau(X)\right) + 2\mathbb{E}\left(X^2\tau(X)\right)\theta.$$

Setting it equal to zero and solving for θ:

$$\theta = \frac{\mathbb{E}\left(YX\tau(X)\right)}{\mathbb{E}\left(X^2\tau(X)\right)}.$$

(ii) Since $e = Y - \theta X$,

$$\begin{aligned}
\mathbb{E}\left(X e \tau(X)\right) &= \mathbb{E}\left(XY\tau(X)\right) - \mathbb{E}\left(X^2\tau(X)\right)\theta \\
&= \mathbb{E}\left(XY\tau(X)\right) - \mathbb{E}\left(X^2\tau(X)\right)\left(\mathbb{E}\left(X^2\tau(X)\right)\right)^{-1}\mathbb{E}\left(XY\tau(X)\right) \\
&= 0.
\end{aligned}$$

(iii)

$$\widehat{\theta} = \frac{\frac{1}{n}\sum_{i=1}^n Y_i X_i \tau(X_i)}{\frac{1}{n}\sum_{i=1}^n X_i^2 \tau(X_i)}.$$

(iv) By WLLN, $\frac{1}{n}\sum_{i=1}^n Y_i X_i \tau(X_i) \to_p \mathbb{E}\left(YX\tau(X)\right)$ and $\frac{1}{n}\sum_{i=1}^n X_i^2 \tau(X_i) \to_p \mathbb{E}\left(X^2\tau(X)\right)$. By continuous mapping theorem, since $(x, y) \mapsto x/y$ is continuous,

$$\widehat{\theta} = \frac{\frac{1}{n}\sum_{i=1}^n Y_i X_i \tau(X_i)}{\frac{1}{n}\sum_{i=1}^n X_i^2 \tau(X_i)} \to_p \frac{\mathbb{E}\left(YX\tau(X)\right)}{\mathbb{E}\left(X^2\tau(X)\right)} = \theta.$$

(v)

$$\widehat{\theta} = \frac{\frac{1}{n}\sum_{i=1}^n Y_i X_i \tau(X_i)}{\frac{1}{n}\sum_{i=1}^n X_i^2 \tau(X_i)} = \frac{\frac{1}{n}\sum_{i=1}^n X_i \tau(X_i)\left(X_i\theta + e_i\right)}{\frac{1}{n}\sum_{i=1}^n X_i^2 \tau(X_i)} = \theta + \frac{\frac{1}{n}\sum_{i=1}^n X_i e_i \tau(X_i)}{\frac{1}{n}\sum_{i=1}^n X_i^2 \tau(X_i)},$$

therefore, since $\mathbb{E}\left(X_i e_i \tau(X_i)\right) = 0$,

$$\begin{aligned}
\sqrt{n}\left(\widehat{\theta} - \theta\right) &= \frac{1}{\frac{1}{n}\sum_{i=1}^n X_i^2 \tau(X_i)}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i e_i \tau(X_i)\right) \\
&\to_d \frac{1}{\mathbb{E}\left(X^2\tau(X)\right)} N\left(0, \mathbb{E}\left(X_i^2 e_i^2 \tau(X_i)^2\right)\right) \sim N\left(0, \frac{\mathbb{E}\left(X_i^2 e_i^2 \tau(X_i)^2\right)}{\mathbb{E}\left(X^2\tau(X)\right)^2}\right).
\end{aligned}$$

Problem 3. (15 Points) Partition the matrix of regressors $\boldsymbol{X}$ as follows:

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

$\boldsymbol{M}_1 = \boldsymbol{I}_n - \boldsymbol{X}_1\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'$. 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:

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

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

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

(iii) 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.

(i) $\widetilde{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{P_X Y} \neq \widehat{\boldsymbol{\beta}}_2$, 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.

(ii) $\widetilde{\boldsymbol{\beta}}_2 = \left(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{Y}$. Therefore the estimates are the same. The residuals from regressing $\boldsymbol{Y}$ on $\boldsymbol{M}_1\boldsymbol{X}_2$ are

$$\boldsymbol{Y} - \boldsymbol{M}_1\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 = \boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 + \widehat{\boldsymbol{e}} - \boldsymbol{M}_1\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 = \boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 + \boldsymbol{P}_1\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 + \widehat{\boldsymbol{e}} = \boldsymbol{P}_1\boldsymbol{Y} + \widehat{\boldsymbol{e}}.$$

The residuals may not be the same as $\widehat{\boldsymbol{e}}$.

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

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

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

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

But in general, $\boldsymbol{P_X Y} \neq \boldsymbol{M}_1\boldsymbol{Y}$.

Problem 4. (10 Points) Let $\boldsymbol{X}$ be the matrix collecting all the n observations on the k regressors. Let $\boldsymbol{Z} = \boldsymbol{XB}$, 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.

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

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

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

Problem 5. (15 Points) Let Y be a random variable and $\boldsymbol{X}$ be a random k-vector. Suppose that the conditional distribution of Y given $\boldsymbol{X}$ is normal: $Y|\boldsymbol{X} \sim N(\boldsymbol{X}'\boldsymbol{\beta}, \sigma^2)$,where $\boldsymbol{\beta}$ is some k-vector and $\sigma^2 > 0$ is some constant.

For a random variable Z, its τ-th quantile is defined by the equation: $\Pr\left(Z \leq z_\tau\right) = \tau$. Similarly, a function $q_\tau(\boldsymbol{X})$ is the τ-th quantile of the conditional distribution of Y given $\boldsymbol{X}$ if $\Pr(Y \leq q_\tau(\boldsymbol{X})|\boldsymbol{X}) = \tau$.

(i) For $\tau \in (0,1)$, find the τ-th quantile for the conditional distribution of Y given $\boldsymbol{X}$. **Hint**: Let z_τ be the τ-th quantile of the standard normal distribution. Consider normalized Y that has a standard normal distribution.

(ii) Compare the marginal effects of X on the conditional mean of Y and on the τ-th conditional quantile of Y. Are they the same or different?

(iii) How will you answer in (ii) change if the model is heteroskedastic: $Y|\boldsymbol{X} \sim N\left(\boldsymbol{X}'\boldsymbol{\beta}, e^{2\boldsymbol{X}'\boldsymbol{\gamma}}\right)$, where $\boldsymbol{\gamma}$ is some k-vector?

Solution.

(i) Let z_τ denote the τ-the quantile of the standard normal distribution. Since

$$\left.\frac{Y - \boldsymbol{X}'\boldsymbol{\beta}}{\sigma}\right| \boldsymbol{X} \sim N(0,1),$$

we have that

$$\begin{aligned} \tau &= \Pr\left(\left.\frac{Y - \boldsymbol{X}'\boldsymbol{\beta}}{\sigma} \le z_\tau \right| \boldsymbol{X}\right) \\ &= \Pr\left(Y \le \sigma z_\tau + \boldsymbol{X}'\boldsymbol{\beta} \middle| \boldsymbol{X}\right). \end{aligned}$$

Thus, the τ-th conditional quantile of Y given $\boldsymbol{X}$ is:

$$q_\tau(\boldsymbol{X}) = \sigma z_\tau + \boldsymbol{X}'\boldsymbol{\beta}.$$

(ii) The marginal effect of X on the τ-th conditional quantile of Y is:

$$\frac{\partial q_\tau(\boldsymbol{X})}{\partial \boldsymbol{X}} = \frac{\partial\left(\sigma z_\tau + \boldsymbol{X}'\boldsymbol{\beta}\right)}{\partial \boldsymbol{X}} = \boldsymbol{\beta}.$$

Since the conditional mean of Y given $\boldsymbol{X}$ is a linear function of $\boldsymbol{X}$, the marginal effect of $\boldsymbol{X}$ on the conditional mean is the same as on the conditional quantile:

$$\frac{\partial \mathbb{E}(Y|\boldsymbol{X})}{\partial \boldsymbol{X}} = \frac{\partial\left(\boldsymbol{X}'\boldsymbol{\beta}\right)}{\partial \boldsymbol{X}} = \boldsymbol{\beta}.$$

(iii) In the heteroskedastic model, by the same argument as in (i), the τ-the conditional quantile of Y is given by

$$q_\tau(\boldsymbol{X}) = z_\tau e^{\boldsymbol{X}'\boldsymbol{\gamma}} + \boldsymbol{X}'\boldsymbol{\beta}.$$

Now,

$$\frac{\partial q_\tau(\boldsymbol{X})}{\partial \boldsymbol{X}} = \frac{\partial\left(z_\tau e^{\boldsymbol{X}'\boldsymbol{\gamma}} + \boldsymbol{X}'\boldsymbol{\beta}\right)}{\partial \boldsymbol{X}} = z_\tau e^{\boldsymbol{X}'\boldsymbol{\gamma}}\boldsymbol{\gamma} + \boldsymbol{\beta},$$

which is different from the marginal effect on the conditional mean if $\boldsymbol{\gamma} \neq \boldsymbol{0}$.

Problem 6. (20 Points) Consider the simple (one-regressor) linear regression model without an intercept: $\boldsymbol{Y} = \beta\boldsymbol{X} + \boldsymbol{e}$, where $\boldsymbol{Y}$, $\boldsymbol{X}$, and $\boldsymbol{e}$ are n-dimensional random vectors, and β is an unknown scalar parameter. Assume that $\mathbb{E}(\boldsymbol{e}|\boldsymbol{X}) = 0$ and $\mathbb{E}(\boldsymbol{e}\boldsymbol{e}'|\boldsymbol{X}) = \sigma^2\boldsymbol{I}_n$.

(i) Show that the LS estimator of β is $\widehat{\beta} = \frac{\boldsymbol{X}'\boldsymbol{Y}}{\boldsymbol{X}'\boldsymbol{X}}$.

(ii) Find $\operatorname{Var}\left(\widehat{\beta}|\boldsymbol{X}\right)$.

(iii) Consider the following estimator of β: $\widetilde{\beta} = \frac{\bar{Y}}{\bar{X}}$, where $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$ and $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Is $\widetilde{\beta}$ unbiased?

(iv) Find $\operatorname{Var}\left(\widetilde{\beta}|\boldsymbol{X}\right)$.

(v) Without relying on Gauss-Markov Theorem, show that $\widetilde{\beta}$ is less efficient than $\widehat{\beta}$. **Hint**: Using Cauchy-Schwartz inequality, show that $\left(\sum_{i=1}^n X_i\right)^2 \leq n\sum_{i=1}^n X_i^2$.

Solution.

(i) Let $\boldsymbol{X} = (X_1, X_2, ..., X_n)'$ and $\boldsymbol{Y} = (Y_1, Y_2, ..., Y_n)'$. Then

$$\widehat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} = \frac{\sum_{i=1}^n X_iY_i}{\sum_{i=1}^n X_i^2}.$$

(ii)

$$\begin{aligned}
\mathrm{Var}(\widehat{\beta}|\boldsymbol{X}) &= \mathrm{Var}((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}|\boldsymbol{X}) \\
&= \mathrm{Var}((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\beta\boldsymbol{X}+\boldsymbol{e})|\boldsymbol{X}) \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathrm{Var}(\boldsymbol{e}|\boldsymbol{X})[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']' \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\sigma^2\boldsymbol{I}_n[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']' \\
&= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= \sigma^2(\sum_{i=1}^n X_i^2)^{-1}.
\end{aligned}$$

(iii)

$$\begin{aligned}
\mathbb{E}(\widetilde{\beta}|\boldsymbol{X}) &= \mathbb{E}(\frac{\beta\bar{X}+\bar{e}}{\bar{X}}|\boldsymbol{X}) \\
&= \beta + \frac{1}{\bar{X}}\mathbb{E}(\bar{e}|\boldsymbol{X}) \\
&= \beta + \frac{1}{\bar{X}}\mathbb{E}(\frac{1}{n}\sum_{i=1}^n e_i|\boldsymbol{X}) \\
&= \beta + \frac{1}{\bar{X}}\frac{1}{n}\sum_{i=1}^n \mathbb{E}(e_i|\boldsymbol{X}) \\
&= \beta.
\end{aligned}$$

So $\widetilde{\beta}$ is unbiased.

(iv)

$$\begin{aligned}
\mathrm{Var}(\widetilde{\beta}|\boldsymbol{X}) &= \mathrm{Var}((\sum_{i=1}^n X_i)^{-1}(\sum_{i=1}^n Y_i)|\boldsymbol{X}) \\
&= \mathrm{Var}((\sum_{i=1}^n X_i)^{-1}(\beta\sum_{i=1}^n X_i + \sum_{i=1}^n e_i)|\boldsymbol{X}) \\
&= (\sum_{i=1}^n X_i)^{-2}\mathrm{Var}(\sum_{i=1}^n e_i|\boldsymbol{X}) \\
&= (\sum_{i=1}^n X_i)^{-2}\sum_{i=1}^n \mathrm{Var}(e_i|\boldsymbol{X}), \text{use } \mathbb{E}(e_ie_j|\boldsymbol{X}) = 0 \text{ implied by } \mathbb{E}(\boldsymbol{e}\boldsymbol{e}'|\boldsymbol{X}) = \sigma^2\boldsymbol{I}_n. \\
&= n(\sum_{i=1}^n X_i)^{-2}\sigma^2.
\end{aligned}$$

(v) Using $(\sum_i^n X_i)^2 = (\sum X_i \cdot 1)^2 \leq \sum X_i^2 \sum_i^n 1^2 = n \sum X_i^2$ which follows from the Cauchy-Schwartz inequality, we can get:

$$(\sum_{i=1}^n X_i^2)^{-1} \leq n(\sum_{i=1}^n X_i)^{-2}.$$

Therefore,

$$\mathrm{Var}(\widehat{\beta}|\boldsymbol{X}) \leq \mathrm{Var}(\widetilde{\beta}|\boldsymbol{X}).$$

Problem 7. (10 Points) Suppose we observe a random sample $\{(Y_i, D_i)\}_{i=1}^n$, where Y_i is the dependent variable and D_i is a binary independent variable: for all $i = 1, 2, ..., n$, $D_i = 1$ or $D_i = 0$. Suppose we regress Y_i on D_i **with an intercept**. Show: the LS estimate of the slope is equal to the difference between the sample averages of the dependent variable of the two groups, observations with $D_i = 1$ and observations with $D_i = 0$. **Hint**: The sample average of Y of observations with $D_i = 1$ can be written as $\frac{\sum_{i=1}^n D_i Y_i}{\sum_{i=1}^n D_i}$. What is the sample average of Y of observations with $D_i = 0$? Also note: $D_i = D_i^2$.

Solution. Denote $\overline{D} = n^{-1} \sum_{i=1}^n D_i$. The LS estimate is

$$\widehat{\beta} = \frac{\sum_{i=1}^n \left(D_i - \overline{D}\right) Y_i}{\sum_{i=1}^n \left(D_i - \overline{D}\right)^2} = \frac{\sum_{i=1} \left(D_i - \overline{D}\right) Y_i}{\sum_{i=1}^n D_i^2 - n\overline{D}^2} = \frac{\sum_{i=1} D_i Y_i - n\overline{DY}}{n\overline{D} - n\overline{D}^2}.$$

The sample average of Y of observations with $D_i = 0$ is

$$\frac{\sum_{i=1}^n (1 - D_i) Y_i}{\sum_{i=1}^n (1 - D_i)}.$$

Then,

$$\begin{aligned}
\frac{\sum_{i=1}^n D_i Y_i}{\sum_{i=1}^n D_i} - \frac{\sum_{i=1}^n (1 - D_i) Y_i}{\sum_{i=1}^n (1 - D_i)} &= \frac{\sum_{i=1}^n D_i Y_i}{n\overline{D}} - \frac{\sum_{i=1}^n (1 - D_i) Y_i}{n - n\overline{D}} \\
&= \frac{\left(n - n\overline{D}\right) \sum_{i=1}^n D_i Y_i - \left(n\overline{D}\right) \sum_{i=1}^n (1 - D_i) Y_i}{n\overline{D}\left(n - n\overline{D}\right)} \\
&= \frac{\sum_{i=1}^n D_i Y_i - \overline{D} \sum_{i=1}^n D_i Y_i - n\overline{DY} + \overline{D} \sum_{i=1}^n D_i Y_i}{n\overline{D} - n\overline{D}^2} \\
&= \widehat{\beta}.
\end{aligned}$$

Problem 8. (**Bonus Question**, 15 Points) Suppose we observe a random sample $\{X_1, ..., X_n\}$ from a uniform distribution on $[0, \theta]$, where $\theta > 0$ is unknown.

(i) Show that $\widehat{\theta} = \max\{X_1, ..., X_n\}$ is a consistent estimator for θ.

(ii) $n\left(\theta - \widehat{\theta}\right)$ converges in distribution to a random variable X that has a exponential distribution:

$$F_X(x) = \Pr[X \leq x] = 1 - e^{-\frac{x}{\theta}}.$$

Hint: $\Pr[\max\{X_1, ..., X_n\} < x] = \Pr[X_1 < x, ..., X_n < x] = \prod_{i=1}^n \Pr[X_i < x]$. Use $\lim_{n\to\infty} \left(1 + \frac{z}{n}\right)^n = e^z$, for all $z \in \mathbb{R}$. Write $\Pr\left[n\left(\theta - \widehat{\theta}\right) \leq x\right] = 1 - \Pr\left[n\left(\theta - \widehat{\theta}\right) > x\right]$ $(x > 0)$ then find the limit.

Solution.

(i) For any (small) $\epsilon > 0$,

$$0 \leq \Pr\left(\left|\widehat{\theta} - \theta\right| \geq \epsilon\right) = \Pr\left(\widehat{\theta} \geq \theta + \epsilon \text{ or } \widehat{\theta} \leq \theta - \epsilon\right)$$

$$
\begin{aligned}
&\leq \Pr\left(\widehat{\theta} \geq \theta + \epsilon\right) + \Pr\left(\widehat{\theta} \leq \theta - \epsilon\right) \\
&= \Pr\left(X_1 \geq \theta + \epsilon \text{ or } X_2 \geq \theta + \epsilon \text{ or } \cdots \text{ or } X_n \geq \theta + \epsilon\right) \\
&\quad + \Pr\left(X_1 \leq \theta - \epsilon, ..., X_n \leq \theta - \epsilon\right) \\
&\leq \sum_{i=1}^{n} \Pr\left(X_i \geq \theta + \epsilon\right) + \prod_{i=1}^{n} \Pr\left(X_i \leq \theta - \epsilon\right) \\
&= 0 + \left(\frac{\theta - \epsilon}{\theta}\right)^n \\
&\to 0 \text{ as } n \to \infty.
\end{aligned}
$$

(ii) For any $x > 0$,

$$
\begin{aligned}
\Pr\left[n\left(\theta - \widehat{\theta}\right) \leq x\right] &= 1 - \Pr\left[n\left(\theta - \widehat{\theta}\right) > x\right] \\
&= 1 - \Pr\left[\theta - \frac{x}{n} > \max\left\{X_1, ..., X_n\right\}\right] \\
&= 1 - \prod_{i=1}^{n} \Pr\left[X_i < \theta - \frac{x}{n}\right] \\
&= 1 - \left(\frac{1}{\theta}\left(\theta - \frac{x}{n}\right)\right)^n \\
&= 1 - \left(1 - \frac{1}{n}\frac{x}{\theta}\right)^n \to 1 - e^{-\frac{x}{\theta}}.
\end{aligned}
$$