Homework 3

Problem 1. For a random variable X, the probability of the event $X \leq c$ (c is a constant) is $\Pr(X \leq c)$. Define $Z = 1(X \leq c)$. Z is a Bernoulli random variable. Prove: $\mathbb{E}Z = \Pr(X \leq c)$.

Solution.

$$\begin{aligned}
\mathbb{E}Z &= 1 \cdot \Pr(Z = 1) + 0 \cdot \Pr(Z = 0) \\
&= 1 \cdot \Pr(X \leq c) + 0 \cdot \Pr(X > 0) \\
&= \Pr(X \leq c).
\end{aligned}$$

Problem 2. Use FWL Theorem to show that in a simple (one-regressor) regression model,

$$Y_i = \beta_0 + \beta_1 X_i + U_i, \ i = 1, \ldots, n,$$

the LS estimate for β_1 is

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n \left(X_i - \overline{X}\right) Y_i}{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2}.$$

Then assume (1) (X_i, Y_i), $i = 1, ..., n$ are independently and identically distributed (i.i.d.). (2) $E(U_i | X_i) = 0$, for $i = 1, ..., n$. (3) $E\left(U_i^2 | X_i\right) = \sigma^2$, for $i = 1, ..., n$, with some $\sigma > 0$. Show that

$$\mathrm{Var}\left(\widehat{\beta}_1 | X_1, ..., X_n\right) = \frac{\sigma^2}{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2}.$$

Solution. $\boldsymbol{M_1} = \boldsymbol{I}_n - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}' = \boldsymbol{I}_n - n^{-1}\mathbf{1}\mathbf{1}'$. Denote $\boldsymbol{X} = (X_1, ..., X_n)'$ and $\boldsymbol{Y} = (Y_1, ..., Y_n)'$. Then, $\boldsymbol{M_1 X} = \boldsymbol{X} - \mathbf{1} \cdot \overline{X}$. By FWL theorem,

$$\begin{aligned}
\widehat{\beta}_1 &= \left(\boldsymbol{X}'\boldsymbol{M_1}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{M_1}\boldsymbol{Y}\right) \\
&= \frac{\left(\boldsymbol{X} - \mathbf{1} \cdot \overline{X}\right)'\boldsymbol{Y}}{\left(\boldsymbol{X} - \mathbf{1} \cdot \overline{X}\right)'\left(\boldsymbol{X} - \mathbf{1} \cdot \overline{X}\right)} \\
&= \frac{\sum_{i=1}^n \left(X_i - \overline{X}\right) Y_i}{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2}.
\end{aligned}$$

Note: $\sum_{i=1}^n \left(X_i - \overline{X}\right) = n \cdot \overline{X} - n \cdot \overline{X} = 0$ and

$$\sum_{i=1}^n \left(X_i - \overline{X}\right) X_i = \sum_{i=1}^n \left(X_i - \overline{X}\right)\left(X_i - \overline{X} + \overline{X}\right) = \sum_{i=1}^n \left(X_i - \overline{X}\right)^2 + \sum_{i=1}^n \left(X_i - \overline{X}\right)\overline{X} = \sum_{i=1}^n \left(X_i - \overline{X}\right)^2.$$

Then,

$$\begin{aligned}
\widehat{\beta}_1 &= \frac{\sum_{i=1}^n \left(X_i - \overline{X}\right)(\beta_0 + \beta_1 X_i + U_i)}{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2} \\
&= \frac{\beta_1 \sum_{i=1}^n \left(X_i - \overline{X}\right) X_i + \sum_{i=1}^n \left(X_i - \overline{X}\right) U_i}{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2} \\
&= \beta_1 + \frac{\sum_{i=1}^n \left(X_i - \overline{X}\right) U_i}{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2}.
\end{aligned}$$

And,

$$
\begin{aligned}
\mathbb{E}\left(\widehat{\beta}_1|\boldsymbol{X}\right) &= \beta_1+\mathbb{E}\left(\sum_{i=1}^{n}\frac{\left(X_i-\overline{X}\right)}{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}U_i\Big|\boldsymbol{X}\right)\\
&= \beta_1+\sum_{i=1}^{n}\frac{\left(X_i-\overline{X}\right)}{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}\mathbb{E}\left(U_i|\boldsymbol{X}\right)\\
&= \beta_1.
\end{aligned}
$$

Then,

$$
\begin{aligned}
\operatorname{Var}\left(\widehat{\beta}_1|\boldsymbol{X}\right) &= \mathbb{E}\left(\left(\widehat{\beta}_1-\mathbb{E}\left(\widehat{\beta}_1|\boldsymbol{X}\right)\right)^2|\boldsymbol{X}\right)\\
&= \mathbb{E}\left(\left(\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)U_i}{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}\right)^2\Big|\boldsymbol{X}\right)\\
&= \frac{1}{\left(\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)^2}\mathbb{E}\left(\left(\sum_{i=1}^{n}\left(X_i-\overline{X}\right)U_i\right)^2|\boldsymbol{X}\right)\\
&= \frac{1}{\left(\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)^2}\left\{\mathbb{E}\left(\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2U_i^2|\boldsymbol{X}\right)+\mathbb{E}\left(\sum_{i\neq j}\left(X_i-\overline{X}\right)\left(X_j-\overline{X}\right)U_iU_j|\boldsymbol{X}\right)\right\}\\
&= \frac{1}{\left(\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)^2}\left\{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\mathbb{E}\left(U_i^2|\boldsymbol{X}\right)+\sum_{i\neq j}\left(X_i-\overline{X}\right)\left(X_j-\overline{X}\right)\mathbb{E}\left(U_iU_j|\boldsymbol{X}\right)\right\}\\
&= \frac{1}{\left(\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)^2}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\sigma^2\\
&= \frac{\sigma^2}{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}.
\end{aligned}
$$

Problem 3. Consider again the simple linear regression model:

$$
Y_i = \beta_0+\beta_1X_i+U_i,\ i=1,\ldots,n;
$$

with assumptions: (1) (X_i,Y_i), $i=1,...,n$ are independently and identically distributed (i.i.d.). (2) $E\left(U_i|X_i\right)=0$, for $i=1,...,n$. (3) $E\left(U_i^2|X_i\right)=\sigma^2$, for $i=1,...,n$, with some $\sigma>0$. Define the estimator

$$
\bar{\beta}_1=\frac{\frac{\sum_{i=1}^{n}Y_i1\{X_i\geq 0\}}{\sum_{i=1}^{n}1\{X_i\geq 0\}}-\frac{\sum_{i=1}^{n}Y_i1\{X_i<0\}}{\sum_{i=1}^{n}1\{X_i<0\}}}{\frac{\sum_{i=1}^{n}X_i1\{X_i\geq 0\}}{\sum_{i=1}^{n}1\{X_i\geq 0\}}-\frac{\sum_{i=1}^{n}X_i1\{X_i<0\}}{\sum_{i=1}^{n}1\{X_i<0\}}}
$$

where

$$
1\left\{X_i\geq 0\right\}=\begin{cases}1 & \text{if } X_i\geq 0\\ 0 & \text{if } X_i<0\end{cases}
$$

and

$$
1\left\{X_i<0\right\}=\begin{cases}1 & \text{if } X_i<0\\ 0 & \text{if } X_i\geq 0.\end{cases}
$$

In other words, $\bar{\beta}_1$ is the difference between the averaged Y's conditional on X being positive and the averaged Y's conditional on X being negative divided by the difference between the averaged X conditional on X being positive and the averaged X conditional on X being negative. Assume $\frac{\sum_{i=1}^n X_i 1\{X_i \geq 0\}}{\sum_{i=1}^n 1\{X_i \geq 0\}} \neq \frac{\sum_{i=1}^n X_i 1\{X_i < 0\}}{\sum_{i=1}^n 1\{X_i < 0\}}$.

(i) Show that $\bar{\beta}_1$ is unbiased.

(ii) Is the conditional variance $\text{Var}\left(\bar{\beta}_1 | X_1, ..., X_n\right)$ less than or equal to $\frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$ (the variance of the LS estimator)? Explain.

Solution. (i) As we have done in class we should: (1) substitute $Y_i = \beta_0 + \beta_1 X_i + U_i$ and then (2) use the properties of expectations to simplify.

$$
\begin{aligned}
E[\bar{\beta}_1] =& E\left[\frac{\frac{\sum_{i=1}^n Y_i 1\{X_i \geq 0\}}{\sum_{i=1}^n 1\{X_i \geq 0\}} - \frac{\sum_{i=1}^n Y_i 1\{X_i < 0\}}{\sum_{i=1}^n 1\{X_i < 0\}}}{\frac{\sum_{i=1}^n X_i 1\{X_i \geq 0\}}{\sum_{i=1}^n 1\{X_i \geq 0\}} - \frac{\sum_{i=1}^n X_i 1\{X_i < 0\}}{\sum_{i=1}^n 1\{X_i < 0\}}}\right] \\
=& E\left[\frac{\frac{\sum_{i=1}^n (\beta_0 + X_i \beta_1 + U_i) 1\{X_i \geq 0\}}{\sum_{i=1}^n 1\{X_i \geq 0\}} - \frac{\sum_{i=1}^n (\beta_0 + X_i \beta_1 + U_i) 1\{X_i < 0\}}{\sum_{i=1}^n 1\{X_i < 0\}}}{\frac{\sum_{i=1}^n X_i 1\{X_i \geq 0\}}{\sum_{i=1}^n 1\{X_i \geq 0\}} - \frac{\sum_{i=1}^n X_i 1\{X_i < 0\}}{\sum_{i=1}^n 1\{X_i < 0\}}}\right] \\
& \text{rearranging} \\
=& E\left[\frac{\left(\beta_0 \frac{\sum_{i=1}^n 1\{X_i \geq 0\}}{\sum_{i=1}^n 1\{X_i \geq 0\}} + \beta_1 \frac{\sum_{i=1}^n X_i 1\{X_i \geq 0\}}{\sum_{i=1}^n 1\{X_i \geq 0\}} + \frac{\sum_{i=1}^n U_i 1\{X_i \geq 0\}}{\sum_{i=1}^n 1\{X_i \geq 0\}}\right) - \left(\beta_0 \frac{\sum_{i=1}^n 1\{X_i < 0\}}{\sum_{i=1}^n 1\{X_i < 0\}} + \beta_1 \frac{\sum_{i=1}^n X_i 1\{X_i < 0\}}{\sum_{i=1}^n 1\{X_i < 0\}} + \frac{\sum_{i=1}^n U_i 1\{X_i < 0\}}{\sum_{i=1}^n 1\{X_i < 0\}}\right)}{\frac{\sum_{i=1}^n X_i 1\{X_i \geq 0\}}{\sum_{i=1}^n 1\{X_i \geq 0\}} - \frac{\sum_{i=1}^n X_i 1\{X_i < 0\}}{\sum_{i=1}^n 1\{X_i < 0\}}}\right] \\
& \text{simplifying} \\
=& \beta_1 + E\left[\frac{\frac{\sum_{i=1}^n U_i 1\{X_i \geq 0\}}{\sum_{i=1}^n 1\{X_i \geq 0\}} - \frac{\sum_{i=1}^n U_i 1\{X_i < 0\}}{\sum_{i=1}^n 1\{X_i < 0\}}}{\frac{\sum_{i=1}^n X_i 1\{X_i \geq 0\}}{\sum_{i=1}^n 1\{X_i \geq 0\}} - \frac{\sum_{i=1}^n X_i 1\{X_i < 0\}}{\sum_{i=1}^n 1\{X_i < 0\}}}\right] \\
& \text{using iterated expectations} \\
=& \beta_1 + E\left[E\left[\frac{\frac{\sum_{i=1}^n U_i 1\{X_i \geq 0\}}{\sum_{i=1}^n 1\{X_i \geq 0\}} - \frac{\sum_{i=1}^n U_i 1\{X_i < 0\}}{\sum_{i=1}^n 1\{X_i < 0\}}}{\frac{\sum_{i=1}^n X_i 1\{X_i \geq 0\}}{\sum_{i=1}^n 1\{X_i \geq 0\}} - \frac{\sum_{i=1}^n X_i 1\{X_i < 0\}}{\sum_{i=1}^n 1\{X_i < 0\}}} | X_1, ..., X_n\right]\right] \\
& \text{using the linearity of } E[\cdot | X_1, ..., X_n] \text{ we have} \\
=& \beta_1 + E\left[\frac{\frac{\sum_{i=1}^n E[U_i | X_1, ..., X_n] 1\{X_i \geq 0\}}{\sum_{i=1}^n 1\{X_i \geq 0\}} - \frac{\sum_{i=1}^n E[U_i | X_1, ..., X_n] 1\{X_i < 0\}}{\sum_{i=1}^n 1\{X_i < 0\}}}{\frac{\sum_{i=1}^n X_i 1\{X_i \geq 0\}}{\sum_{i=1}^n 1\{X_i \geq 0\}} - \frac{\sum_{i=1}^n X_i 1\{X_i < 0\}}{\sum_{i=1}^n 1\{X_i < 0\}}}\right] \\
& E[U_i | X_1, ..., X_n] = 0 \text{ by assumption, so} \\
=& \beta_1
\end{aligned}
$$

(ii) The previous part showed $\bar{\beta}_1$ is unbiased. It is also linear because it is equal $\sum_{i=1}^n \bar{c}_i Y_i$ with

$$
\bar{c}_i = \frac{\frac{1\{X_i \geq 0\}}{\sum_{i=1}^n 1\{X_i \geq 0\}} - \frac{1\{X_i < 0\}}{\sum_{i=1}^n 1\{X_i < 0\}}}{\frac{\sum_{i=1}^n X_i 1\{X_i \geq 0\}}{\sum_{i=1}^n 1\{X_i \geq 0\}} - \frac{\sum_{i=1}^n X_i 1\{X_i < 0\}}{\sum_{i=1}^n 1\{X_i < 0\}}}.
$$

Therefore, by the Gauss-Markov theorem, $Var(\bar{\beta}_1) > Var(\hat{\beta}_1)$.

Problem 4. Suppose we observe an i.i.d. random sample $\{X_1, X_2, ..., X_n\}$ with $\mathbb{E}X_1^2 < \infty$. Denote $\mu = \mathbb{E}X_1$ and $\sigma^2 = \text{Var}(X_1)$. (a) Give an unbiased but inconsistent estimator for μ; (b) Give a consistent but biased estimator for μ.

Solution. (a) Unbiased but inconsistent: X_1. $\mathbb{E}X_1 = \mu$, but for any $\epsilon > 0$, $\Pr(|X_1 - \mu| \leq \epsilon)$ is constant:

$$
\lim_{n \to \infty} \Pr(|X_1 - \mu| \leq \epsilon) = \Pr(|X_1 - \mu| \leq \epsilon) \neq 0.
$$

(b) Consistent but biased: $(n-1)^{-1} \sum_{i=1}^n X_i$. $\mathbb{E}\left((n-1)^{-1} \sum_{i=1}^n X_i\right) = (n/(n-1)) \mu \neq \mu$.

$$
\frac{1}{n-1} \sum_{i=1}^n X_i = \frac{n}{n-1} \frac{1}{n} \sum_{i=1}^n X_i \to \mu, \text{ as } n \to \infty.
$$

Problem 5. (a) Prove the "Squeeze Rule": If $0 \leq X_n \leq Y_n$ and $Y_n \to_p 0$, then $X_n \to_p 0$; (b) Prove: $X_n \to_p 0$ if and only if $|X_n| \to_p 0$.

Solution. For any $\epsilon > 0$,

$$\Pr(Y_n \leq \epsilon) \leq \Pr(X_n \leq \epsilon) \leq 1.$$

Then,

$$\Pr(Y_n \leq \epsilon) = \Pr(|Y_n - 0| \leq \epsilon) \to 1 \Longrightarrow \Pr(X_n \leq \epsilon) = \Pr(|X_n - 0| \leq \epsilon) \to 1.$$

Problem 6. Provide a counter example to show that $X_n \to_d X$ and $Y_n \to_d Y$ does not imply $X_n + Y_n \to_d X + Y$. Hint: Consider an iid random sample $X_1, ..., X_n$ with $\mathbb{E}X_1 = 0$ and $n^{1/2}\overline{X}_n$ and $-n^{1/2}\overline{X}_n$.

Solution. Let Z be a random variable such that $Z \sim N\left(0, \sigma^2\right)$, where $\sigma^2 = \mathrm{Var}(X_1)$. Then by CLT,

$$n^{1/2}\overline{X}_n \to_d Z$$

and

$$-n^{1/2}\overline{X}_n = (-1) \times \left(n^{1/2}\overline{X}_n\right) \to_d (-1) \times Z \sim N\left(0, \sigma^2\right).$$

Therefore, it is also true that $-n^{1/2}\overline{X}_n \to_d Z \sim N\left(0, \sigma^2\right)$. Note

$$0 = \left(n^{1/2}\overline{X}_n\right) + \left(-n^{1/2}\overline{X}_n\right) \not\to_d Z + Z \sim N\left(0, 4\sigma^2\right).$$

Problem 7. Let $\widehat{\boldsymbol{\theta}}_n = \left(\widehat{\theta}_{n,1}, \ldots, \widehat{\theta}_{n,k}\right)'$ be an estimator of the k-vector of parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$. Suppose that $\widehat{\boldsymbol{\theta}}_n \to_p \boldsymbol{\theta}$, and $n^{1/2}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \to_d \boldsymbol{W} \sim N(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a positive definite $k \times k$ matrix. Use the delta method or CMT to find the (non-degenerate, i.e., not a constant) asymptotic distributions of the following quantities after a suitable normalization. "Suitable normalization" means subtraction of a constant and/or multiplication by a constant (could be dependent on n).

(i) $n^{1/2}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)'\boldsymbol{c}$ where $\boldsymbol{c} \in \mathbb{R}^k$ is a vector of constants.

(ii) $\widehat{\theta}_{n,1}$.

(iii) $n\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)'\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)$.

(iv) $\widehat{\theta}_{n,1} - \widehat{\theta}_{n,2}$.

(v) $\widehat{\theta}_{n,1}\widehat{\theta}_{n,2}/\widehat{\theta}_{n,3}$, provided that $\theta_3 \neq 0$.

Solution.

(i) Define $\boldsymbol{X}_n = n^{1/2}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)$ and $h(\boldsymbol{X}_n) = \boldsymbol{X}_n'\boldsymbol{c}$. By the Continuous Mapping Theorem we have

$$n^{1/2}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)'\boldsymbol{c} = h(\boldsymbol{X}_n) \to_d h(\boldsymbol{W}) = \boldsymbol{W}'\boldsymbol{c}$$

By the property of normal distribution we have

$$n^{1/2}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)'\boldsymbol{c} \to_d \boldsymbol{W}'\boldsymbol{c} \sim N(\boldsymbol{0}, \boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c}).$$

(ii) Set $\boldsymbol{c} = (1, 0, ..., 0)'$. Then, it follows from Part (i) that

$$n^{1/2}(\widehat{\theta}_{n,1} - \theta_1) \to_d N(0, \sigma_{11}^2),$$

where σ_{11}^2 is the first diagonal element of $\boldsymbol{\Sigma}$.

(iii) Since $n^{1/2}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \to_d \boldsymbol{W}$, by the Continuous Mapping Theorem,

$$n(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) = \left[n^{1/2}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)\right]' \left[n^{1/2}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)\right] \to_d \boldsymbol{W}'\boldsymbol{W}.$$

(iv) Set $\boldsymbol{c} = (1, -1, 0, ..., 0)'$.It follows from Part (i) that

$$n^{1/2}(\widehat{\theta}_{n,1} - \widehat{\theta}_{n,2} - \theta_1 + \theta_2) \to_d N(0, \sigma_{11}^2 - 2\sigma_{12} + \sigma_{22}^2),$$

where σ_{11}^2 and σ_{22}^2 are the first and second diagonal element of $\boldsymbol{\Sigma}$, and σ_{12} is the element on the first row and second column of $\boldsymbol{\Sigma}$.

(v) Put $h(\boldsymbol{\theta}) = \frac{\theta_1\theta_2}{\theta_3}$, apply the Delta method

$$n^{1/2}(\frac{\widehat{\theta}_{n,1}\widehat{\theta}_{n,2}}{\widehat{\theta}_{n,3}} - \frac{\theta_1\theta_2}{\theta_3}) = n^{1/2}(h(\widehat{\boldsymbol{\theta}}_n) - h(\boldsymbol{\theta})) \to_d \frac{\partial h(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\boldsymbol{W}$$

where

$$\frac{\partial h(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \left(\frac{\theta_2}{\theta_3}, \frac{\theta_1}{\theta_3}, \frac{-\theta_1\theta_2}{\theta_3^2}, 0, ..., 0\right)'.$$

Then by the property of Normal density

$$n^{1/2}(\frac{\widehat{\theta}_{n,1}\widehat{\theta}_{n,2}}{\widehat{\theta}_{n,3}} - \frac{\theta_1\theta_2}{\theta_3}) \to_d N(0, \frac{\partial h(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\boldsymbol{\Sigma}\frac{\partial h(\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}}).$$

Problem 8. Suppose that $\hat{\theta}_n \to_p \theta$ and $\hat{\beta}_n \to \beta$, where θ and β are two scalar parameters. Without relying on Slutsky's Theorem, show:

(i) $c\hat{\theta}_n \to_p c\theta$, where c is a constant.

(ii) $\hat{\theta}_n\hat{\beta}_n \to_p \theta\beta$.

Solution. (i) Suppose $c \neq 0$. Then $\Pr\left(|c\hat{\theta}_n - c\theta| > \varepsilon\right) = \Pr\left(|\hat{\theta}_n - \theta| > \frac{\varepsilon}{|c|}\right) \to 0$ as $n \to \infty$. If $c = 0$, then $c\hat{\theta}_n = 0 \to_p c\theta = 0$.

(ii) First, note that $\hat{\theta}_n\hat{\beta}_n - \theta\beta = (\hat{\theta}_n - \theta + \theta)(\hat{\beta}_n - \beta + \beta) - \theta\beta = \left(\hat{\theta}_n - \theta\right)\left(\hat{\beta}_n - \beta\right) + \left(\hat{\theta}_n - \theta\right)\beta + \left(\hat{\beta}_n - \beta\right)\theta$. Then, $\left(\hat{\theta}_n - \theta\right)\beta + \left(\hat{\beta}_n - \beta\right)\theta \to_p 0$ by Part (i). Then, for any $\epsilon > 0$,

$$\begin{aligned} \Pr\left(\left|\left(\hat{\theta}_n - \theta\right)\left(\hat{\beta}_n - \beta\right)\right| > \varepsilon\right) &\leq \Pr\left(\left|\hat{\theta}_n - \theta\right| > \sqrt{\varepsilon} \text{ or } \left|\hat{\beta}_n - \beta\right| > \sqrt{\varepsilon}\right) \\ &\leq \Pr\left(\left|\hat{\theta}_n - \theta\right| > \sqrt{\varepsilon}\right) + \Pr\left(\left|\hat{\beta}_n - \beta\right| > \sqrt{\varepsilon}\right) \to 0 \text{ as } n \to \infty. \end{aligned}$$

Thus, $\hat{\theta}_n\hat{\beta}_n - \theta\beta \to_p 0$.

Problem 9. Suppose that $\mathbb{E}\left(\hat{\theta}_n\right) \to \theta$ and $\text{Var}(\hat{\theta}_n) \to 0$ as $n \to \infty$. Show that $\hat{\theta}_n \to_p \theta$.

Solution. $\hat{\theta}_n$ converges in probability to θ if for all $\varepsilon > 0$, $\Pr\left(\left|\hat{\theta}_n - \theta\right| \geq \varepsilon\right) \to 0$ as $n \to \infty$. First, decompose the Mean Squared Error (MSE) into

$$MSE\left(\hat{\theta}_n\right) = \mathbb{E}\left(\hat{\theta}_n - \theta\right)^2 = \mathbb{E}\left(\hat{\theta}_n - \mathbb{E}\hat{\theta}_n + \mathbb{E}\hat{\theta}_n - \theta\right)^2$$

$$= \mathbb{E}\left(\hat{\theta}_n - \mathbb{E}\hat{\theta}_n\right)^2 + \left(\mathbb{E}\hat{\theta}_n - \theta\right)^2 + 2\mathbb{E}\left(\hat{\theta}_n - \mathbb{E}\hat{\theta}_n\right)\left(\mathbb{E}\hat{\theta}_n - \theta\right)$$
$$= \mathbb{E}\left(\hat{\theta}_n - \mathbb{E}\hat{\theta}_n\right)^2 + \left(\mathbb{E}\hat{\theta}_n - \theta\right)^2 = \mathrm{Var}\left(\hat{\theta}_n\right) + \mathrm{Bias}\left(\hat{\theta}_n\right)^2,$$

where the last line follows by the fact that $\mathbb{E}\left(\hat{\theta}_n - \mathbb{E}\hat{\theta}_n\right) = 0$.

Then, using Markov's Inequality,

$$\Pr\left(\left|\hat{\theta}_n - \theta\right| \geq \varepsilon\right) \leq \frac{\mathbb{E}\left|\hat{\theta}_n - \theta\right|^2}{\varepsilon^2} = \frac{\mathbb{E}\left(\hat{\theta}_n - \theta\right)^2}{\varepsilon^2} = \frac{\mathrm{Var}\left(\hat{\theta}_n\right) + \mathrm{Bias}\left(\hat{\theta}_n\right)^2}{\varepsilon^2} \to 0 \text{ as } n \to \infty,$$

since by assumption, $\mathrm{Var}\left(\hat{\theta}_n\right) \to 0$ and $\mathbb{E}\hat{\theta}_n - \theta \to 0$ as $n \to \infty$.

Problem 10. 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 $\mathbb{E}\left(\boldsymbol{e}|\boldsymbol{X}\right) = 0$ and $\mathbb{E}\left(\boldsymbol{e}\boldsymbol{e}'|\boldsymbol{X}\right) = \sigma^2 \boldsymbol{I}_n$.

(i) Show that the LS estimator of β is

$$\widehat{\beta} = \frac{\boldsymbol{X}'\boldsymbol{Y}}{\boldsymbol{X}'\boldsymbol{X}}.$$

(ii) Define the fitted residuals $\widehat{\boldsymbol{e}} = \boldsymbol{Y} - \widehat{\beta}\boldsymbol{X}$. $\boldsymbol{Y} = (Y_1, ..., Y_n)'$, $\boldsymbol{X} = (X_1, ..., X_n)'$, $\boldsymbol{e} = (e_1, ..., e_n)'$, $\widehat{\boldsymbol{e}} = (\widehat{e}_1, ..., \widehat{e}_n)'$. For each of the following statements, explain if it is true or false:

(a) $\mathbb{E}(e_i X_i) = 0$ for all $i = 1, \ldots, n$.
(b) $\mathbb{E}e_i = 0$ for all $i = 1, \ldots, n$.
(c) $\sum_{i=1}^n \widehat{e}_i X_i = 0$.
(d) $\sum_{i=1}^n \widehat{e}_i = 0$.
(e) $\sum_{i=1}^n e_i X_i = 0$.
(f) $\sum_{i=1}^n e_i = 0$.

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

(iv) Consider the following estimator of β:

$$\widetilde{\beta} = \frac{\bar{Y}}{\bar{X}},$$

where

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i \text{ and } \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i.$$

Is $\widetilde{\beta}$ unbiased?

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

(vi) 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

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

(ii) (a) True. By the law of iterated expectations:

$$\mathbb{E}\,(e_i X_i) = \mathbb{E}\mathbb{E}(e_i X_i|\boldsymbol{X}) = \mathbb{E}\mathbb{E}(e_i|\boldsymbol{X})X_i = 0$$

by the assumption that $\mathbb{E}(\boldsymbol{e}|\boldsymbol{X}) = \mathbf{0}$. (b) True. By the law of iterated expectations:

$$\mathbb{E}e_i = \mathbb{E}\mathbb{E}(e_i|\boldsymbol{X}) = 0$$

by the assumption that $\mathbb{E}(\boldsymbol{e}|\boldsymbol{X}) = \mathbf{0}$. (c) True.

$$\sum_{i=1}^n \widehat{e}_i X_i = \sum_{i=1}^n (Y_i - \widehat{\beta}X_i)X_i = \sum_{i=1}^n (Y_i X_i - \widehat{\beta}X_i^2) = 0.$$

(d) False. If $\sum_{i=1}^n \widehat{e}_i = 0$, $\widehat{\beta}$ must also solve

$$\sum_{i=1}^n Y_i - \widehat{\beta}\sum_{i=1}^n X_i = 0 \quad (2^*)$$

i.e.

$$\widehat{\beta} = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i}.$$

Unless equations (1^*) and (2^*) are linearly dependent, $\widehat{\beta} = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}$ cannot solve them simultaneously. (e) False. $\sum_{i=1}^n e_i X_i$ is a sum of limited realizations of a random variable and thus the sum itself is a random variable, it will be zero with probability zero. Thus in general we do not have $\sum_{i=1}^n e_i X_i = \mathbb{E}(\sum_{i=1}^n e_i X_i) = 0$. (f) False. Same reason.

(iii)

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

(iv)

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

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

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

(v)

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

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

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

Therefore,

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

Problem 11. Consider the following model:

$$
Y_i = \beta + U_i,
$$

where U_i are iid $N(0,1)$ random variables, $i = 1,\ldots,n$.

(i) Find the LS estimator of β and its mean, variance, and distribution.

(ii) Suppose that a data set of 100 observation produced OLS estimate $\widehat{\beta} = 0.167$.

- (a) Construct 90% and 95% symmetric two-sided confidence intervals for β.
- (b) Construct a 95% one-sided confidence interval of the form $[A,+\infty)$ for β. In other words, find a random variable A such that $\Pr(\beta \in [A,+\infty)) = 1-\alpha$, where $\alpha \in (0, 0.5)$ is a known constant chosen by the econometrician.
- (c) Construct a 95% one-sided confidence interval of the form $(-\infty, A]$ for β.

Solution. The model is $Y_i = \beta + U_i$, with $\{U_i\}_{i=1}^n$ i.i.d random variables and $U_i \sim N(0,1)$, $i = 1,...,n$.

LS estimator for β is given by $\widehat{\beta} = (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{Y}$, where $\mathbf{1}$ is a $n\times 1$ vector of ones and $\boldsymbol{Y} = (Y_1,\cdots,Y_n)'$. Therefore, $\widehat{\beta} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \overline{Y}$. Notice the following

$$
\widehat{\beta} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n}(\beta + U_i) = \beta + \frac{1}{n}\sum_{i=1}^{n} U_i.
$$

Hence,

$$\begin{aligned}
\mathbb{E}\widehat{\beta} &= \beta + \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(U_i) = \beta \\
\mathrm{Var}(\widehat{\beta}) &= \mathrm{Var}(\frac{1}{n}\sum_{i=1}^{n}U_i) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(U_i) \quad \text{since } U_i\text{'s are i.i.d.} \\
&= \frac{n}{n^2} = \frac{1}{n}.
\end{aligned}$$

Since $\widehat{\beta}$ is just a linear combination of iid normal random variables, $\widehat{\beta} \sim N(\beta, \frac{1}{n})$. $\widehat{\beta} = 0.167$. Confidence interval for significance level α is

$$\widehat{\beta} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \leq \beta \leq \widehat{\beta} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}$$

Plugging in the values for $\widehat{\beta} = 0.167$, $\sqrt{\frac{\sigma^2}{n}} = 0.1$, $z_{1-\frac{\alpha}{2}} = 1.645$ when $\alpha = 0.1$, $z_{1-\frac{\alpha}{2}} = 1.96$ when $\alpha = 0.05$. We obtain $CI_{90\%} = [0.0025, 0.3315]$ and $CI_{95\%} = [-0.029, 0.363]$.

One sided confidence interval for significance level $\alpha = 0.05$ of the form $[a, +\infty)$ is

$$\beta \geq \widehat{\beta} - z_{1-\alpha}\sqrt{\frac{\sigma^2}{n}}.$$

Plugging in the values for $\widehat{\beta} = 0.167$, $\sqrt{\frac{\sigma^2}{n}} = 0.1$, $z_{1-\alpha} = 1.645$. We obtain the one-sided confidence interval $CI_{95\%} = [0.0025, \infty)$.

One sided confidence interval for significance level $\alpha = 0.05$ of the form $(-\infty, a]$ is

$$\beta \leq \hat{\beta} + z_{1-\alpha}\sqrt{\frac{\sigma^2}{n}}$$

Plugging in the values for $\widehat{\beta} = 0.167$, $\sqrt{\frac{\sigma^2}{n}} = 0.1$, $z_{1-\alpha} = 1.645$. We obtain the one-sided confidence interval $CI_{95\%} = (-\infty, 0.3315]$.

Problem 12. 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.

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

(ii) Define

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

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

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

(iv) Assume that

$$\mathbb{E}\left(\boldsymbol{V}\boldsymbol{V}'|\boldsymbol{X}_1\right) = \sigma_v^2 I_n,$$

and find $\text{Var}(\tilde{\boldsymbol{\beta}}_1|\boldsymbol{X}_1)$.

(v) 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 $\text{Var}(\tilde{\boldsymbol{\beta}}_1|\boldsymbol{X}_1)$ derived in part (iv) with $\text{Var}(\hat{\boldsymbol{\beta}}_1|\boldsymbol{X}_1, \boldsymbol{X}_2)$. Can you say which of the two variances is bigger (in the positive semi-definite sense)? Explain your answer.

Solution. The LS estimator for $\boldsymbol{\beta}_1$ which omits $\boldsymbol{\beta}_2$ from the regression is $\widetilde{\boldsymbol{\beta}}_1 = \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{Y}$ that can be written as

$$\widetilde{\boldsymbol{\beta}}_1 = \boldsymbol{\beta}_1 + \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\boldsymbol{\beta}_2 + \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{e}.$$

(i)

$$\mathbb{E}\left(\widetilde{\boldsymbol{\beta}}_1|\boldsymbol{X}_1\right) = \boldsymbol{\beta}_1 + \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\mathbb{E}\left(\boldsymbol{X}_2\boldsymbol{\beta}_2|\boldsymbol{X}_1\right) + \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\mathbb{E}\left(\boldsymbol{e}|\boldsymbol{X}_1\right).$$

By Law of Iterated Expectations, $\mathbb{E}\left(\boldsymbol{e}|\boldsymbol{X}_1\right) = \mathbb{E}\left(\mathbb{E}\left(\boldsymbol{e}|\boldsymbol{X}_1, \boldsymbol{X}_2\right)|\boldsymbol{X}_1\right) = \boldsymbol{0}$, thus

$$\mathbb{E}\left(\widetilde{\boldsymbol{\beta}}_1|\boldsymbol{X}_1\right) = \boldsymbol{\beta}_1 + \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\mathbb{E}\left(\boldsymbol{X}_2\boldsymbol{\beta}_2|\boldsymbol{X}_1\right).$$

Also, by defining $\boldsymbol{V}$ as $\boldsymbol{V} = \boldsymbol{X}_2\boldsymbol{\beta}_2 - \mathbb{E}\left(\boldsymbol{X}_2\boldsymbol{\beta}_2|\boldsymbol{X}_1\right)$, $\widetilde{\boldsymbol{\beta}}_1 - \mathbb{E}\left(\widetilde{\boldsymbol{\beta}}_1|\boldsymbol{X}_1\right) = \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{V} + \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{e}$.

(ii) In order to find $\mathbb{E}\left(\boldsymbol{e}\boldsymbol{V}'|\boldsymbol{X}_1\right) = \mathbb{E}\left[\boldsymbol{e}\left(\boldsymbol{X}_2\boldsymbol{\beta}_2 - \mathbb{E}\left(\boldsymbol{X}_2\boldsymbol{\beta}_2|\boldsymbol{X}_1\right)\right)'|\boldsymbol{X}_1\right]$, use again the Law of Iterated Expectations,

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{e}\left(\boldsymbol{X}_2\boldsymbol{\beta}_2 - \mathbb{E}\left(\boldsymbol{X}_2\boldsymbol{\beta}_2|\boldsymbol{X}_1\right)\right)'|\boldsymbol{X}_1\right] &= \mathbb{E}\left(\mathbb{E}\left[\boldsymbol{e}\left(\boldsymbol{X}_2\boldsymbol{\beta}_2 - \mathbb{E}\left(\boldsymbol{X}_2\boldsymbol{\beta}_2|\boldsymbol{X}_1\right)\right)'|\boldsymbol{X}_1, \boldsymbol{X}_2\right]|\boldsymbol{X}_1\right) \\
&= \mathbb{E}\left(\mathbb{E}\left[\boldsymbol{e}|\boldsymbol{X}_1, \boldsymbol{X}_2\right]\left(\boldsymbol{X}_2\boldsymbol{\beta}_2 - \mathbb{E}\left(\boldsymbol{X}_2\boldsymbol{\beta}_2|\boldsymbol{X}_1\right)\right)'|\boldsymbol{X}_1\right) \\
&= \mathbb{E}\left(\boldsymbol{0}\left(\boldsymbol{X}_2\boldsymbol{\beta}_2 - \mathbb{E}\left(\boldsymbol{X}_2\boldsymbol{\beta}_2|\boldsymbol{X}_1\right)\right)'|\boldsymbol{X}_1\right) \\
&= \boldsymbol{0}.
\end{aligned}$$

(iii)

$$\mathbb{E}\left(\boldsymbol{e}\boldsymbol{e}'|\boldsymbol{X}_1\right) = \mathbb{E}\left(\mathbb{E}\left(\boldsymbol{e}\boldsymbol{e}'|\boldsymbol{X}_1, \boldsymbol{X}_2\right)|\boldsymbol{X}_1\right) = \mathbb{E}\left(\sigma_e^2\boldsymbol{I}_n|\boldsymbol{X}_1\right) = \sigma_e^2\boldsymbol{I}_n.$$

(iv) Using previous results and the fact that $\mathbb{E}\left(\boldsymbol{V}\boldsymbol{V}'|\boldsymbol{X}_1\right) = \sigma_v^2\boldsymbol{I}_n$,

$$\begin{aligned}
\text{Var}\left(\widetilde{\boldsymbol{\beta}}_1|\boldsymbol{X}_1\right) &= \mathbb{E}\left(\left[\widetilde{\boldsymbol{\beta}}_1 - E\left(\widetilde{\boldsymbol{\beta}}_1|\boldsymbol{X}_1\right)\right]\left[\widetilde{\boldsymbol{\beta}}_1 - E\left(\widetilde{\boldsymbol{\beta}}_1|\boldsymbol{X}_1\right)\right]'|\boldsymbol{X}_1\right) \\
&= \mathbb{E}\left(\left[\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{V} + \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{e}\right]\left[\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{V} + \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{e}\right]'|\boldsymbol{X}_1\right) \\
&= \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\mathbb{E}\left(\left(\boldsymbol{e}+\boldsymbol{V}\right)\left(\boldsymbol{e}+\boldsymbol{V}\right)'|\boldsymbol{X}_1\right)\boldsymbol{X}_1\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1} \\
&= \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\mathbb{E}\left(\boldsymbol{e}\boldsymbol{e}' + \boldsymbol{e}\boldsymbol{V}' + \boldsymbol{V}\boldsymbol{e}' + \boldsymbol{V}\boldsymbol{V}'|\boldsymbol{X}_1\right)\boldsymbol{X}_1\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1} \\
&= \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\left(\sigma_e^2\boldsymbol{I}_n + \sigma_v^2\boldsymbol{I}_n\right)\boldsymbol{X}_1\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1} \\
&= \left(\sigma_e^2 + \sigma_v^2\right)\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}.
\end{aligned}$$

(v) $\text{Var}\left(\widehat{\beta}_1|\boldsymbol{X}_1\right) = \sigma_e^2\left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)^{-1}$. Then, $\boldsymbol{X}_1'\boldsymbol{X}_1 - \boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1 = \boldsymbol{X}_1'\boldsymbol{P}_2\boldsymbol{X}_1 \geq 0$ since $\boldsymbol{P}_2$ is a projection matrix (symmetric and idempotent), therefore positive semi-definite. It follows that $\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1} - \left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)^{-1} \leq 0$. Therefore, since $\sigma_v^2 > 0$, it is ambiguous which variance is larger, $\text{Var}\left(\widetilde{\boldsymbol{\beta}}_1|\boldsymbol{X}_1\right)$ or $\text{Var}\left(\widehat{\beta}_1|\boldsymbol{X}_1\right)$.

Problem 13. Let $X_1, \dots, X_n$ be random variables with finite variances. Show that

$$\begin{aligned}\operatorname{Var}\left(\sum_{i=1}^n X_i\right) &= \sum_{i=1}^n \sum_{j=1}^n \operatorname{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n \operatorname{Var}(X_i) + \sum_{i=1}^n \sum_{j \neq i} \operatorname{Cov}(X_i, X_j).\end{aligned}$$

Hint: $\sum_{i=1}^n X_i = \mathbf{1}' (X_1, ..., X_n)$.

Solution. Denote $\boldsymbol{X} = (X_1, ..., X_n)'$. Then,

$$\begin{aligned}\operatorname{Var}\left(\sum_{i=1}^n X_i\right) &= \operatorname{Var}(\mathbf{1}'\boldsymbol{X}) = \mathbf{1}'\operatorname{Var}(\boldsymbol{X})\,\mathbf{1} = \mathbf{1}'\begin{pmatrix} \operatorname{Cov}(X_1, X_1) & \operatorname{Cov}(X_1, X_2) & \cdots & \operatorname{Cov}(X_1, X_n) \\ \operatorname{Cov}(X_2, X_1) & \operatorname{Cov}(X_2, X_2) & \cdots & \operatorname{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{Cov}(X_n, X_1) & \operatorname{Cov}(X_n, X_2) & \cdots & \operatorname{Cov}(X_n, X_n) \end{pmatrix}\mathbf{1} \\ &= \sum_{i=1}^n \sum_{j=1}^n \operatorname{Cov}(X_i, X_j).\end{aligned}$$

Problem 14. 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: $\Pr(Z \leq z_\tau) = \tau$. 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?

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

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

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:

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