Homework 3

Problem 1. 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 2. Suppose that assumptions of the Classical Linear Regression model hold, i.e.

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

however,

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

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

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

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

(in the positive semidefinite sense). Hint: Show

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

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

Solution.

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

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

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

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

3. Following the hint we first show,

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

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

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

Problem 3. 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 $S(\boldsymbol{b})$.

Solution.

1.

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

2. By adding and subtracting $\boldsymbol{X}\widetilde{\boldsymbol{\beta}}$, we have

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

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

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

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

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

Problem 4. 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) $E\left(U_i|X_i\right)=0$, for $i=1,...,n$. (3) $E\left(U_i^2|X_i\right)=\sigma^2$, for $i=1,...,n$, with some $\sigma>0$. Show that

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

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

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

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,

$$\begin{aligned} \widehat{\beta}_1 &= \frac{\sum_{i=1}^n\left(X_i-\overline{X}\right)\left(\beta_0+\beta_1X_i+U_i\right)}{\sum_{i=1}^n\left(X_i-\overline{X}\right)^2} \\ &= \frac{\beta_1\sum_{i=1}^n\left(X_i-\overline{X}\right)X_i+\sum_{i=1}^n\left(X_i-\overline{X}\right)U_i}{\sum_{i=1}^n\left(X_i-\overline{X}\right)^2} \\ &= \beta_1+\frac{\sum_{i=1}^n\left(X_i-\overline{X}\right)U_i}{\sum_{i=1}^n\left(X_i-\overline{X}\right)^2}. \end{aligned}$$

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,

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

Problem 5. Consider again the simple linear regression model:

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

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

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

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 $\frac{\sum_{i=1}^n X_i 1\{X_i \geq 0\}}{\sum_{i=1}^n 1\{X_i \geq 0\}} \neq \frac{\sum_{i=1}^n X_i 1\{X_i < 0\}}{\sum_{i=1}^n 1\{X_i < 0\}}$.

1. Show that $\bar{\beta}_1$ is unbiased.
2. Is the conditional variance $\text{Var}\left(\bar{\beta}_1 | X_1, ..., X_n\right)$ less than or equal to $\frac{\sigma^2}{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}$ (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 6. Suppose that a random variable X has a normal distribution with unknown mean μ. To simplify the analysis, we shall assume that σ^2 is known. Given a sample of observations, an estimator of μ is the sample mean, $\overline{X}$. When performing a (two-sided) test of the null hypothesis $H_0 : \mu = \mu_0$ at 5% significance level, it is usual to choose the upper and lower 2.5% tails of the normal distribution as the rejection regions, as shown in the first figure. s.d. is equal to $\sqrt{\sigma^2/n}$, the standard deviation of $\overline{X}$. The density function of $N\left(\mu_0, \sigma^2/n\right)$ is shown in the first figure. H_0 is rejected when $\left|\overline{X} - \mu_0\right|/\text{s.d.} > 1.96$. However, suppose that someone instead chooses the central 5% of the distribution as the rejection region, as in the second figure. Give a technical explanation, using appropriate statistical concepts, of why this is not a good idea.

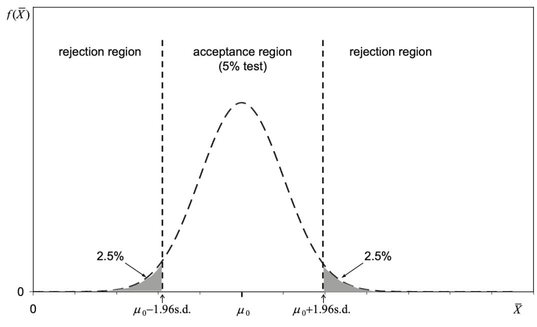

Figure 1: Conventional rejection regions.

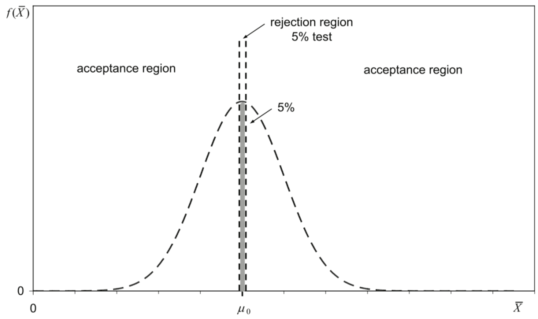

Figure 2: Central 5 per cent chosen as rejection region.

Solution. The following discussion assumes that you are performing a 5 per cent significance test, but it applies to any significance level. If the null hypothesis is true, it does not matter how you define the 5 per cent rejection region. By construction, the risk of making a Type I error will be 5 per cent. Issues relating to Type II errors are irrelevant when the null hypothesis is true.

The reason that the central part of the conditional distribution is not used as a rejection region is that it leads to problems when the null hypothesis is false. The probability of not rejecting H_0 when it is false will be lower. To use the obvious technical term, the power of the test will be lower. The figure opposite shows the power functions for the test using the conventional upper and lower 2.5 per cent tails and the test using the central region. The horizontal axis is the difference between the true value and the hypothetical value μ_0 in terms of standard deviations. The vertical axis is the power of the test. The first figure has been drawn for the case where the true value is greater than the hypothetical value. The second figure is for the case where the true value is lower than the hypothetical value. It is the same, but reflected horizontally.

The greater the difference between the true value and the hypothetical mean, the more likely it is that the sample mean will lie in a tail of the distribution conditional on H_0 being true, and so the more likely it is that the null hypothesis will be rejected by the conventional test. The figures show that the power of the test approaches 1 asymptotically. However, if the central region of the distribution is used as the rejection region, the probability of the sample mean lying in it will diminish as the difference between the true and hypothetical values increases, and the power of the test approaches zero asymptotically. This is an extreme example of a very bad test procedure.

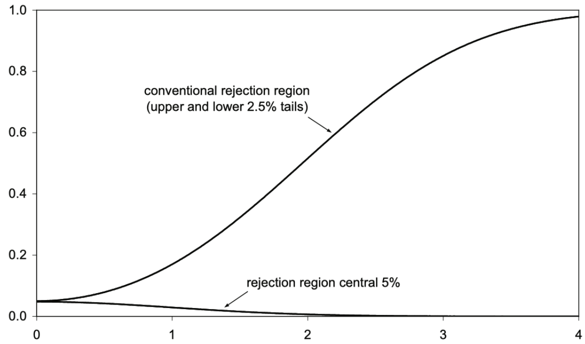

Figure 3: Power functions of a conventional 5 per cent test and one using the central region (true value $> \mu_0$).

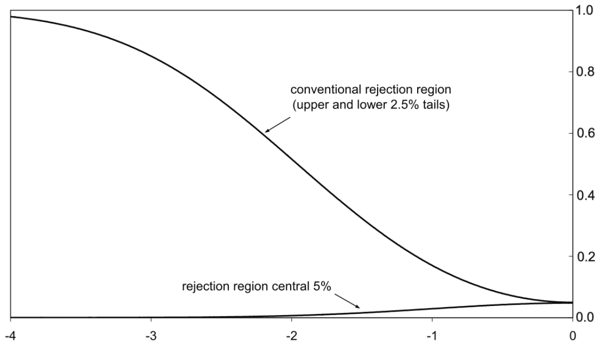

Figure 4: Power functions of a conventional 5 per cent test and one using the central region (true value $< \mu_0$).

Problem 7. Consider the following model:

$$Y_i = \beta + U_i,$$

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

1. Find the LS estimator of β and its mean, variance, and distribution.
2. 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,

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

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

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

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

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

2. 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,
$$\begin{aligned}\mathbb{E}\left[\boldsymbol{e}\left(\boldsymbol{X}_2\boldsymbol{\beta}_2 - \mathbb{E}\left(\boldsymbol{X}_2\boldsymbol{\beta}_2|\boldsymbol{X}_1\right)\right)'|\boldsymbol{X}_1\right] &= \mathbb{E}\left(\mathbb{E}\left[\boldsymbol{e}\left(\boldsymbol{X}_2\boldsymbol{\beta}_2 - \mathbb{E}\left(\boldsymbol{X}_2\boldsymbol{\beta}_2|\boldsymbol{X}_1\right)\right)'|\boldsymbol{X}_1,\boldsymbol{X}_2\right]|\boldsymbol{X}_1\right)\\ &= \mathbb{E}\left(\mathbb{E}\left[\boldsymbol{e}|\boldsymbol{X}_1,\boldsymbol{X}_2\right]\left(\boldsymbol{X}_2\boldsymbol{\beta}_2 - \mathbb{E}\left(\boldsymbol{X}_2\boldsymbol{\beta}_2|\boldsymbol{X}_1\right)\right)'|\boldsymbol{X}_1\right)\\ &= \mathbb{E}\left(\boldsymbol{0}\left(\boldsymbol{X}_2\boldsymbol{\beta}_2 - \mathbb{E}\left(\boldsymbol{X}_2\boldsymbol{\beta}_2|\boldsymbol{X}_1\right)\right)'|\boldsymbol{X}_1\right)\\ &= \boldsymbol{0}.\end{aligned}$$

3.
$$\mathbb{E}\left(\boldsymbol{e}\boldsymbol{e}'|\boldsymbol{X}_1\right) = \mathbb{E}\left(\mathbb{E}\left(\boldsymbol{e}\boldsymbol{e}'|\boldsymbol{X}_1,\boldsymbol{X}_2\right)|\boldsymbol{X}_1\right) = \mathbb{E}\left(\sigma_e^2\boldsymbol{I}_n|\boldsymbol{X}_1\right) = \sigma_e^2\boldsymbol{I}_n.$$

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

5. $\text{Var}\left(\widehat{\beta}_1|\boldsymbol{X}_1\right) = \sigma_e^2\left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)^{-1}$. 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 $\text{Var}\left(\widehat{\beta}_1|\boldsymbol{X}_1\right)$.