Multiple Linear Regression

William Brasic

The University of Arizona

Goal of Multiple Linear Regression (MLR)

Question 1

What is the goal of multiple linear regression (MLR)?

Goal of Multiple Linear Regression (MLR)

Question 1

What is the goal of multiple linear regression (MLR)?

Answer to Question 1

1. Reduce potential covariates excluded from model that are correlated with included regressors.
2. Predict outcome better.
3. Analyze more complex relationships.

MLR Example

Example 1: MLR Example

If we are interested in determinants of a student's GPA, our MLR model could be

$$gpa = \beta_0 + \beta_1 hours\_studied + \beta_2 hours\_studied^2 + \beta_3 male + \beta_4 age + \beta_5 sat\_score + \beta_6 parent\_education + u.$$

- What is the effect of how many hours you spend studying on your GPA?

Multiple Linear Regression

Definition 1: Multiple Linear Regression Using Vectors

The multiple linear regression model in vector notation is given by

$$y_i = \boldsymbol{x}_i' \boldsymbol{\beta} + u_i$$

for $i = 1, \ldots, n$.

- y_i is observation i's outcome.
- $\boldsymbol{x}_i$ is individual i's $(k+1) \times 1$ vector of covariates.
- $\boldsymbol{\beta}$ is the $(k+1) \times 1$ vector of population parameters.
- u_i is observation i's idiosyncratic shock component.

Multiple Linear Regression

Definition 2: Multiple Linear Regression Using Matrices

The multiple linear regression model in matrix notation is given by

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{u}.$$

- $\boldsymbol{y}$ is the $n \times 1$ vector of outcomes for all individuals.
- X is the $n \times (k+1)$ feature matrix.
- $\boldsymbol{\beta}$ is the $(k+1) \times 1$ vector of population parameters.
- $\boldsymbol{u}$ is the $n \times 1$ vector of errors for all individuals.

Multiple Linear Regression

Definition 3: Estimated Multiple Linear Regression

The estimated multiple linear regression model in matrix notation is given by

$$\widehat{\boldsymbol{y}} = X\widehat{\boldsymbol{\beta}}.$$

- $\widehat{\boldsymbol{y}}$ is the $n \times 1$ vector of fitted (prediction) values for all individuals.
- X is the $n \times (k+1)$ feature matrix.
- $\widehat{\boldsymbol{\beta}}$ is the $(k+1) \times 1$ vector of estimated parameters.

Multiple Linear Regression

Definition 4: Multiple Linear Regression Statistics

All definitions discussed for simple linear regression remain the same expect we can represent them nicely using vectors and matrices:

1. $\widehat{\boldsymbol{y}} = X\widehat{\boldsymbol{\beta}}$.
2. $\widehat{\boldsymbol{u}} = \boldsymbol{y} - \widehat{\boldsymbol{y}}$.
3. $SSR = \widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}}$.
4. $SST = \left(\boldsymbol{y} - \overline{Y}\right)'\left(\boldsymbol{y} - \overline{Y}\right)$.
5. $SSE = \left(\widehat{\boldsymbol{y}} - \overline{Y}\right)'\left(\widehat{\boldsymbol{y}} - \overline{Y}\right)$.
6. $\widetilde{R}^2 = 1 - \frac{SSR/(n-k)}{SST/(n-1)}$ where k is the number of estimated parameters (intercept included).

The OLS Solution

Theorem 1: OLS Solution

The solution, $\widehat{\boldsymbol{\beta}}$, to the OLS problem is given by

$$\widehat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y}.$$

- We proved this in the Matrix Algebra lecture slides.

Linear in Parameters

MLR Assumption 1: Linear in Parameters

The population model is a linear function of the parameters.

- For instance, $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{u}$.

Random Sampling

MLR Assumption 2: Random Sampling

We have a random (i.i.d.) sample $\{(y_i, \boldsymbol{x}_i)\}_{i=1}^n$ from the population of interest.

- This will ensure Assumption 4 holds for the entire sample and not just subsets.

No Perfect Multicollinearity

MLR Assumption 3: No Perfect Multicollinearity

No Perfect Multicollinearity means no regressor in the matrix X is constant, nor is any regressor a perfect linear combination of other regressors.

- Also called the full rank condition
  - Implies that $X'X$ is invertible, ensuring that our OLS solution exists and is unique.

Zero Conditional Mean

MLR Assumption 4: Zero Conditional Mean

The expectation of the error term conditioned on the regressors is zero, i.e., $\mathbb{E}[\boldsymbol{u} \mid X] = \mathbf{0}$.

- Also called the exogeneity assumption.
- Important implications include:
  1. $\mathbb{E}[\boldsymbol{u}] = \mathbb{E}[\mathbb{E}[\boldsymbol{u} \mid X]] = \mathbb{E}[\mathbf{0}] = \mathbf{0}$.
  2. $\text{Cov}[X, \boldsymbol{u}] = \mathbb{E}[X'\boldsymbol{u}] - \mathbb{E}[X']\mathbb{E}[\boldsymbol{u}] = \mathbb{E}[X'\boldsymbol{u}] = \mathbb{E}[\mathbb{E}[X'\boldsymbol{u} \mid X]] = \mathbb{E}[X\mathbb{E}[\boldsymbol{u} \mid X]] = \mathbf{0}$.
  3. $\mathbb{E}[\boldsymbol{y} \mid X] = \mathbb{E}[x\boldsymbol{\beta} + \boldsymbol{u} \mid X] = X\boldsymbol{\beta}$ so $\boldsymbol{\beta}$ represents the average impact of our covariates on y_i.

Unbiasedness of OLS for MLR

Theorem 2: Unbiasedness of OLS for MLR

Under MLR Assumptions 1-4, the OLS estimator is unbiased, i.e.,

$$\mathbb{E}\left[\widehat{\boldsymbol{\beta}} \mid X\right] = \boldsymbol{\beta}.$$

- By the law of total expectation $\mathbb{E}\left[\widehat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta}$.
- On average, our estimator is equal to the truth
  - The sampling distribution of our estimated parameters is centered around their true values.

Unbiasedness of OLS for MLR

Proof 1: Unbiased of OLS for MLR Part 1

Using MLR Assumption 1 of $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{u}$ and MLR Assumption 3 that X has full rank so $(X'X)^{-1}$ exists,

$$\widehat{\boldsymbol{\beta}} = (X'X)^{-1} X'\boldsymbol{y}$$

Unbiasedness of OLS for MLR

Proof 1: Unbiased of OLS for MLR Part 1

Using MLR Assumption 1 of $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{u}$ and MLR Assumption 3 that X has full rank so $(X'X)^{-1}$ exists,

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= (X'X)^{-1} X'\boldsymbol{y} \\ &= (X'X)^{-1} X'(X\boldsymbol{\beta} + \boldsymbol{u})\end{aligned}$$

Unbiasedness of OLS for MLR

Proof 1: Unbiased of OLS for MLR Part 1

Using MLR Assumption 1 of $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{u}$ and MLR Assumption 3 that X has full rank so $(X'X)^{-1}$ exists,

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= (X'X)^{-1} X'\boldsymbol{y} \\ &= (X'X)^{-1} X' (X\boldsymbol{\beta} + \boldsymbol{u}) \\ &= (X'X)^{-1} X'X\boldsymbol{\beta} + (X'X)^{-1} X'\boldsymbol{u}\end{aligned}$$

Unbiasedness of OLS for MLR

Proof 1: Unbiased of OLS for MLR Part 1

Using MLR Assumption 1 of $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{u}$ and MLR Assumption 3 that X has full rank so $(X'X)^{-1}$ exists,

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}} &= (X'X)^{-1} X'\boldsymbol{y} \\
&= (X'X)^{-1} X' (X\boldsymbol{\beta} + \boldsymbol{u}) \\
&= (X'X)^{-1} X'X\boldsymbol{\beta} + (X'X)^{-1} X'\boldsymbol{u} \\
&= \boldsymbol{\beta} + (X'X)^{-1} X'\boldsymbol{u}.
\end{aligned}
$$

Unbiasedness of OLS for MLR

Proof 1: Unbiased of OLS for MLR Part 2

Using MLR Assumption 2 of random sampling implying each individual's conditional error term has identical mean and MLR Assumption 4 so this conditional mean is zero,

$$\mathbb{E}\left[\widehat{\boldsymbol{\beta}} \mid X\right] = \mathbb{E}\left[\boldsymbol{\beta} + \left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right]$$

Unbiasedness of OLS for MLR

Proof 1: Unbiased of OLS for MLR Part 2

Using MLR Assumption 2 of random sampling implying each individual's conditional error term has identical mean and MLR Assumption 4 so this conditional mean is zero,

$$\begin{aligned}\mathbb{E}\left[\widehat{\boldsymbol{\beta}} \mid X\right] &= \mathbb{E}\left[\boldsymbol{\beta} + \left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right] \\ &= \boldsymbol{\beta} + \left(X'X\right)^{-1} X'\mathbb{E}\left[\boldsymbol{u} \mid X\right]\end{aligned}$$

Unbiasedness of OLS for MLR

Proof 1: Unbiased of OLS for MLR Part 2

Using MLR Assumption 2 of random sampling implying each individual's conditional error term has identical mean and MLR Assumption 4 so this conditional mean is zero,

$$\begin{aligned}\mathbb{E}\left[\widehat{\boldsymbol{\beta}} \mid X\right] &= \mathbb{E}\left[\boldsymbol{\beta} + \left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right] \\ &= \boldsymbol{\beta} + \left(X'X\right)^{-1} X'\mathbb{E}\left[\boldsymbol{u} \mid X\right] \\ &= \boldsymbol{\beta} + \left(X'X\right)^{-1} X'\mathbf{0}\end{aligned}$$

Unbiasedness of OLS for MLR

Proof 1: Unbiased of OLS for MLR Part 2

Using MLR Assumption 2 of random sampling implying each individual's conditional error term has identical mean and MLR Assumption 4 so this conditional mean is zero,

$$\begin{aligned}\mathbb{E}\left[\widehat{\boldsymbol{\beta}} \mid X\right] &= \mathbb{E}\left[\boldsymbol{\beta} + \left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right] \\ &= \boldsymbol{\beta} + \left(X'X\right)^{-1} X'\mathbb{E}\left[\boldsymbol{u} \mid X\right] \\ &= \boldsymbol{\beta} + \left(X'X\right)^{-1} X'\mathbf{0} \\ &= \boldsymbol{\beta}. \quad \square\end{aligned}$$

Hooray!

Omitted Variable Bias (OVB)

Definition 5: Omitted Variable Bias (OVB)

The omitted variable bias (OVB) problem is when we exclude a covariate from the model (so it is contained in u_i) that is correlated with one of the included covariates as well as the outcome.

Omitted Variable Bias (OVB)

Example 2: Omitted Variable Bias (OVB)

Suppose the true model is $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$, but we only regress y_i on x_i giving the equation $\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1}$. The estimate of $\widehat{\beta}_1$ will be biased when the covariance between x_{i1} and x_{i2} is non-zero.

- Excluding relevant regressors could hurt us.
- Including irrelevant regressors does not hurt us, but makes our estimator less efficient.

Omitted Variable Bias (OVB)

TABLE 3.2 Summary of Bias in $\tilde{\beta}_1$ When x_2 Is Omitted in Estimating Equation (3.40)

	$\text{Corr}(x_1, x_2) > 0$	$\text{Corr}(x_1, x_2) < 0$
$\beta_2 > 0$	Positive bias	Negative bias
$\beta_2 < 0$	Negative bias	Positive bias

- For instance, if x_2 has a positive effect on y ($\beta_2 > 0$) and x_1 and x_2 are positively correlated, then $\widehat{\beta}_1$ will be inflated.
- Bias is multiplicative.

Homoskedasticity

MLR Assumption 5: Homoskedastic Errors

Homoskedasticity states $\mathbb{V}[u_i \mid \boldsymbol{x}_i] = \sigma^2$ for each $i = 1, \ldots, n$. Using vectors, $\mathbb{V}[\boldsymbol{u}|X] = \sigma^2 I_n$.

- Under MLR Assumptions 1-5, $\mathbb{V}[u_i] = \mathbb{E}[\mathbb{V}[u_i \mid \boldsymbol{x}_i]] + \mathbb{V}[\mathbb{E}[u_i \mid \boldsymbol{x}_i]] = \mathbb{E}[\sigma^2] = \sigma^2$.
- The variance of our error is constant across all observations when conditioning on our regressors.
- When this assumption fails, we say the errors are heteroskedastic.
  - Really not that important since White's 1980 correction.

Variance-Covariance Matrix of the OLS Estimator

Theorem 3: Variance-Covariance Matrix of the OLS Estimator

Under MLR Assumptions 1-5, the variance-covariance matrix of the OLS estimator is given by

$$\mathbb{V}\left[\widehat{\boldsymbol{\beta}} \mid X\right] = \sigma^2 \left(X'X\right)^{-1}.$$

- By the law of total variance, $\mathbb{V}\left[\widehat{\boldsymbol{\beta}}\right] = \sigma^2 \left(X'X\right)^{-1}$.

Variance of OLS Estimator

Proof 2: Variance-Covariance Matrix of the OLS Estimator

$$\mathbb{V}\left[\widehat{\boldsymbol{\beta}} \mid X\right]$$
$$= \mathbb{V}\left[\boldsymbol{\beta} + \left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right]$$

Variance of OLS Estimator

Proof 2: Variance-Covariance Matrix of the OLS Estimator

$$\mathbb{V}\left[\widehat{\boldsymbol{\beta}} \mid X\right]$$

$$= \mathbb{V}\left[\boldsymbol{\beta} + (X'X)^{-1} X'\boldsymbol{u} \mid X\right]$$

$$= \mathbb{V}\left[\boldsymbol{\beta} \mid X\right] + \mathbb{V}\left[(X'X)^{-1} X'\boldsymbol{u} \mid X\right] + 2\text{Cov}\left[\boldsymbol{\beta}, (X'X)^{-1} X'\boldsymbol{u} \mid X\right]$$

Variance of OLS Estimator

Proof 2: Variance-Covariance Matrix of the OLS Estimator

$$
\begin{aligned}
&\mathbb{V}\left[\widehat{\boldsymbol{\beta}} \mid X\right] \\
&= \mathbb{V}\left[\boldsymbol{\beta} + (X'X)^{-1} X'\boldsymbol{u} \mid X\right] \\
&= \mathbb{V}\left[\boldsymbol{\beta} \mid X\right] + \mathbb{V}\left[(X'X)^{-1} X'\boldsymbol{u} \mid X\right] + 2\text{Cov}\left[\boldsymbol{\beta}, (X'X)^{-1} X'\boldsymbol{u} \mid X\right] \\
&= \mathbb{V}\left[(X'X)^{-1} X'\boldsymbol{u} \mid X\right]
\end{aligned}
$$

Variance of OLS Estimator

Proof 2: Variance-Covariance Matrix of the OLS Estimator

$$
\begin{aligned}
&\mathbb{V}\left[\widehat{\boldsymbol{\beta}} \mid X\right] \\
&= \mathbb{V}\left[\boldsymbol{\beta} + \left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right] \\
&= \mathbb{V}\left[\boldsymbol{\beta} \mid X\right] + \mathbb{V}\left[\left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right] + 2\text{Cov}\left[\boldsymbol{\beta}, \left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right] \\
&= \mathbb{V}\left[\left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right] \\
&= \left(X'X\right)^{-1} X'\mathbb{V}\left[\boldsymbol{u} \mid X\right] X \left(X'X\right)^{-1}
\end{aligned}
$$

Variance of OLS Estimator

Proof 2: Variance-Covariance Matrix of the OLS Estimator

$$\begin{aligned}
&\mathbb{V}\left[\widehat{\boldsymbol{\beta}} \mid X\right] \\
&= \mathbb{V}\left[\boldsymbol{\beta} + \left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right] \\
&= \mathbb{V}\left[\boldsymbol{\beta} \mid X\right] + \mathbb{V}\left[\left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right] + 2\text{Cov}\left[\boldsymbol{\beta}, \left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right] \\
&= \mathbb{V}\left[\left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right] \\
&= \left(X'X\right)^{-1} X'\mathbb{V}\left[\boldsymbol{u} \mid X\right] X \left(X'X\right)^{-1} \\
&= \left(X'X\right)^{-1} X'\sigma^2 I_n X \left(X'X\right)^{-1}
\end{aligned}$$

Variance of OLS Estimator

Proof 2: Variance-Covariance Matrix of the OLS Estimator

$$
\begin{aligned}
&\mathbb{V}\left[\widehat{\boldsymbol{\beta}} \mid X\right] \\
&= \mathbb{V}\left[\boldsymbol{\beta} + \left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right] \\
&= \mathbb{V}\left[\boldsymbol{\beta} \mid X\right] + \mathbb{V}\left[\left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right] + 2\text{Cov}\left[\boldsymbol{\beta}, \left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right] \\
&= \mathbb{V}\left[\left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right] \\
&= \left(X'X\right)^{-1} X'\mathbb{V}\left[\boldsymbol{u} \mid X\right] X \left(X'X\right)^{-1} \\
&= \left(X'X\right)^{-1} X'\sigma^2 I_n X \left(X'X\right)^{-1} \\
&= \sigma^2 \left(X'X\right)^{-1} X'X \left(X'X\right)^{-1}
\end{aligned}
$$

Variance of OLS Estimator

Proof 2: Variance-Covariance Matrix of the OLS Estimator

$$
\begin{aligned}
&\mathbb{V}\left[\widehat{\boldsymbol{\beta}} \mid X\right] \\
&= \mathbb{V}\left[\boldsymbol{\beta} + \left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right] \\
&= \mathbb{V}\left[\boldsymbol{\beta} \mid X\right] + \mathbb{V}\left[\left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right] + 2\text{Cov}\left[\boldsymbol{\beta}, \left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right] \\
&= \mathbb{V}\left[\left(X'X\right)^{-1} X'\boldsymbol{u} \mid X\right] \\
&= \left(X'X\right)^{-1} X'\mathbb{V}\left[\boldsymbol{u} \mid X\right] X \left(X'X\right)^{-1} \\
&= \left(X'X\right)^{-1} X'\sigma^2 I_n X \left(X'X\right)^{-1} \\
&= \sigma^2 \left(X'X\right)^{-1} X'X \left(X'X\right)^{-1} \\
&= \sigma^2 \left(X'X\right)^{-1}. \quad \square
\end{aligned}
$$

Hooray!

Estimator of the Variance of the OLS Estimator

Theorem 4: Estimator of the Variance-Covariance Matrix of the OLS Estimator

Under MLR Assumption 1-5, since $\widehat{\sigma}^2 = \frac{\widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}}}{n-k}$ is an unbiased estimator for σ^2,

$$\widehat{\mathbb{V}}\left[\widehat{\boldsymbol{\beta}} \mid X\right] = \widehat{\sigma}^2 \left(X'X\right)^{-1}$$

is unbiased estimator for $\mathbb{V}\left[\widehat{\boldsymbol{\beta}} \mid X\right]$.

- k is the number of regressors in our model (including the intercept).

Standard Errors

Definition 6: Standard Errors of OLS Estimates

The standard errors of our OLS estimators are the square roots of the elements along the diagonal of the estimated variance-covariance matrix $\widehat{\mathbb{V}}\left[\widehat{\beta} \mid X\right] = \widehat{\sigma}^2(X'X)^{-1}$.

- The standard error of an estimator is simply its estimated standard deviation.

Why Do We Like OLS?

Question 2

Why do we like OLS so much?

Why Do We Like OLS?

Question 2

Why do we like OLS so much?

Answer to Question 2

There are plenty of unbiased estimators, but under Assumptions 1-5 OLS is efficient, i.e.,

$$\mathbb{V}\left[\widehat{\boldsymbol{\beta}} \mid X\right] \leq \mathbb{V}\left[\widetilde{\boldsymbol{\beta}} \mid X\right]$$

for all other unbiased estimators $\widetilde{\boldsymbol{\beta}}$.

BLUE

Property 1: OLS is BLUE

Under MLR Assumptions 1-5, OLS is BLUE:

1. Best
   - Smallest variance of any other unbiased estimator
2. Linear
3. Unbiased
4. Estimator

- The Gauss-Markov Theorem proves this.

The Gauss-Markov Theorem

Theorem 5: The Gauss-Markov Theorem

Under MLR Assumptions 1-5, OLS is BLUE meaning

$$\mathbb{V}\left[\widehat{\boldsymbol{\beta}}\right] \leq \mathbb{V}\left[\widetilde{\boldsymbol{\beta}}\right]$$

for all other unbiased estimators $\widetilde{\boldsymbol{\beta}}$.

The Gauss-Markov Theorem

Proof 3: The Gauss-Markov Theorem Part 1

Let $\widetilde{\boldsymbol{\beta}}$ be another linear and unbiased estimator of $\boldsymbol{\beta}$ for the model $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{u}$. This implies we can write this estimator as $\widetilde{\boldsymbol{\beta}} = C\boldsymbol{y}$. Furthermore, under MLR Assumption 4 it is necessarily unbiased because $\mathbb{E}\left[\widetilde{\boldsymbol{\beta}}\right] = C\mathbb{E}[\boldsymbol{y}] = CX\boldsymbol{\beta}$ which holds only when $CX = I$.

The Gauss-Markov Theorem

Proof 3: The Gauss-Markov Theorem Part 2

Then,

$$\mathbb{V}\left[\widetilde{\boldsymbol{\beta}}\right] = \mathbb{V}\left[C\boldsymbol{y}\right]$$

The Gauss-Markov Theorem

Proof 3: The Gauss-Markov Theorem Part 2

Then,

$$\begin{aligned}\mathbb{V}\left[\widetilde{\boldsymbol{\beta}}\right] &= \mathbb{V}\left[C\boldsymbol{y}\right] \\ &= C\mathbb{V}\left[\boldsymbol{y}\right]C'\end{aligned}$$

The Gauss-Markov Theorem

Proof 3: The Gauss-Markov Theorem Part 2

Then,

$$\begin{aligned}\mathbb{V}\left[\widetilde{\boldsymbol{\beta}}\right] &= \mathbb{V}\left[C\boldsymbol{y}\right] \\ &= C\mathbb{V}\left[\boldsymbol{y}\right]C' \\ &= \sigma^2 CC'\end{aligned}$$

The Gauss-Markov Theorem

Proof 3: The Gauss-Markov Theorem Part 2

Then,

$$\begin{aligned}\mathbb{V}\left[\tilde{\boldsymbol{\beta}}\right] &= \mathbb{V}\left[C\boldsymbol{y}\right] \\ &= C\mathbb{V}\left[\boldsymbol{y}\right]C' \\ &= \sigma^2 CC' \\ &\geq \sigma^2 CP_X C'\end{aligned}$$

The Gauss-Markov Theorem

Proof 3: The Gauss-Markov Theorem Part 2

Then,

$$\begin{aligned} \mathbb{V}\left[\widetilde{\boldsymbol{\beta}}\right] &= \mathbb{V}\left[C\boldsymbol{y}\right] \\ &= C\mathbb{V}\left[\boldsymbol{y}\right]C' \\ &= \sigma^2 CC' \\ &\geq \sigma^2 CP_X C' \\ &= \sigma^2 CX(X'X)^{-1}X'C' \end{aligned}$$

The Gauss-Markov Theorem

Proof 3: The Gauss-Markov Theorem Part 2

Then,

$$\begin{aligned}
\mathbb{V}\left[\widetilde{\boldsymbol{\beta}}\right] &= \mathbb{V}\left[C\boldsymbol{y}\right] \\
&= C\mathbb{V}\left[\boldsymbol{y}\right]C' \\
&= \sigma^2 CC' \\
&\geq \sigma^2 CP_X C' \\
&= \sigma^2 CX(X'X)^{-1}X'C' \\
&= \sigma^2 (X'X)^{-1}
\end{aligned}$$

The Gauss-Markov Theorem

Proof 3: The Gauss-Markov Theorem Part 2

Then,

$$\begin{aligned}
\mathbb{V}\left[\widetilde{\boldsymbol{\beta}}\right] &= \mathbb{V}\left[C\boldsymbol{y}\right] \\
&= C\mathbb{V}\left[\boldsymbol{y}\right]C' \\
&= \sigma^2 CC' \\
&\geq \sigma^2 C P_X C' \\
&= \sigma^2 CX(X'X)^{-1}X'C' \\
&= \sigma^2 (X'X)^{-1} \\
&= \mathbb{V}\left[\widehat{\boldsymbol{\beta}}\right].
\end{aligned}$$

The Gauss-Markov Theorem

Proof 3: The Gauss-Markov Theorem Part 2

Then,

$$\begin{aligned} \mathbb{V}\left[\widetilde{\boldsymbol{\beta}}\right] &= \mathbb{V}\left[C\boldsymbol{y}\right] \\ &= C\mathbb{V}\left[\boldsymbol{y}\right]C' \\ &= \sigma^2 CC' \\ &\geq \sigma^2 CP_X C' \\ &= \sigma^2 CX(X'X)^{-1}X'C' \\ &= \sigma^2 (X'X)^{-1} \\ &= \mathbb{V}\left[\widehat{\boldsymbol{\beta}}\right]. \end{aligned}$$

Thus, $\mathbb{V}\left[\widetilde{\boldsymbol{\beta}}\right] \geq \mathbb{V}\left[\widehat{\boldsymbol{\beta}}\right]$.

The Gauss-Markov Theorem

Proof 3: The Gauss-Markov Theorem Part 3

Showing $\sigma^2 CC' \geq \sigma^2 CP_X C'$ amounts to showing $\sigma^2 CC' - \sigma^2 CP_X C' \geq 0$, i.e., this matrix is positive semi-definite. A matrix A is positive semi-definite if for any non-zero vector $\boldsymbol{v}$ it follows that $\boldsymbol{v}' A \boldsymbol{v} \geq 0$. Define $A = I - P_X = I - X(X'X)^{-1}X'$. Then,

$$\begin{aligned}
\boldsymbol{v}'(\sigma^2 CC' - \sigma^2 CP_X C')\boldsymbol{v} &= \boldsymbol{v}'(CAC')\boldsymbol{v} \\
&= (C'\boldsymbol{v})' A (C'\boldsymbol{v}) \\
&= (C'\boldsymbol{v})' A' A (C'\boldsymbol{v}) \\
&= (AC'\boldsymbol{v})'(AC'\boldsymbol{v}) \\
&\geq 0
\end{aligned}$$

where $(AC'\boldsymbol{v})'(AC'\boldsymbol{v}) \geq 0$ because the dot product between two identical vectors must be at least zero. □

Thank You!