Lecture 5

September 6, 2018

Last class

- Recall the least squares estimator defined by $\hat{\beta} = (X^T X)^{-1} X^T Y$.

Last class

- Recall the least squares estimator defined by $\hat{\beta} = (X^T X)^{-1} X^T Y$.
- This estimates β in the linear model $Y = X\beta + e$ where $\mathbb{E}(e|X) = 0$.

Last class

- Recall the least squares estimator defined by $\hat{\beta} = (X^T X)^{-1} X^T Y$.
- This estimates β in the linear model $Y = X\beta + e$ where $\mathbb{E}(e|X) = 0$.
- We saw in the last class that $\hat{\beta}$ is linear in Y and conditionally unbiased i.e $\mathbb{E}(\hat{\beta}) = \beta$.

Last class

- Recall the least squares estimator defined by $\hat{\beta} = (X^T X)^{-1} X^T Y$.
- This estimates β in the linear model $Y = X\beta + e$ where $\mathbb{E}(e|X) = 0$.
- We saw in the last class that $\hat{\beta}$ is linear in Y and conditionally unbiased i.e $\mathbb{E}(\hat{\beta}) = \beta$.
- Under the additional assumption of homoskedasticity i.e., $Cov(e|X) = \sigma^2 I_n$, we also calculated the covariance matrix of $\hat{\beta}$ to be $Cov(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$.

Gauss-Markov Theorem and Proof

- The Gauss-Markov Theorem states that $\hat{\beta}$ is BLUE (Best Linear Unbiased Estimator) under homoskedasticity.

Gauss-Markov Theorem and Proof

$\boxed{\text{Four key words, cannot delete anyone}}$

- The Gauss-Markov Theorem states that $\hat{\beta}$ is BLUE (Best Linear Unbiased Estimator) under homoskedasticity.
- This means that $\hat{\beta}$ is the "best" estimator among all linear and unbiased estimators of β.

Gauss-Markov Theorem and Proof

- The Gauss-Markov Theorem states that $\hat{\beta}$ is BLUE (Best Linear Unbiased Estimator) under homoskedasticity.
- This means that $\hat{\beta}$ is the “best" estimator among all linear and unbiased estimators of β.
- Here “best" is in terms of variance. This implies that $\hat{\beta}_i$ has the smallest variance among all linear and unbiased estimators of β_i for every i.

Gauss-Markov Theorem and Proof

- The Gauss-Markov Theorem states that $\hat{\beta}$ is BLUE (Best Linear Unbiased Estimator) under homoskedasticity.
- This means that $\hat{\beta}$ is the “best" estimator among all linear and unbiased estimators of β.
- Here “best" is in terms of variance. This implies that $\hat{\beta}_i$ has the smallest variance among all linear and unbiased estimators of β_i for every i.
- The Gauss-Markov theorem is actually quite simple to prove. Suppose $\tilde{\beta} = AY$ is any other linear unbiased estimator for β.

Gauss-Markov Theorem and Proof

- We get back the least squares estimator if $A = (X^T X)^{-1} X^T$ so let us assume that $B := A - (X^T X)^{-1} X^T \neq 0$.

Gauss-Markov Theorem and Proof

- We get back the least squares estimator if $A = (X^T X)^{-1} X^T$ so let us assume that $B := A - (X^T X)^{-1} X^T \neq 0$.
- Now the expectation of $\tilde{\beta}$ equals $\beta + BX\beta$.

Gauss-Markov Theorem and Proof

- We get back the least squares estimator if $A = (X^T X)^{-1} X^T$ so let us assume that $B := A - (X^T X)^{-1} X^T \neq 0$.
- Now the expectation of $\tilde{\beta}$ equals $\beta + BX\beta$.
- Because $\tilde{\beta}$ is unbiased, this must mean that $BX = 0$.

Gauss-Markov Theorem and Proof

- We get back the least squares estimator if $A = (X^T X)^{-1} X^T$ so let us assume that $B := A - (X^T X)^{-1} X^T \neq 0$.
- Now the expectation of $\tilde{\beta}$ equals $\beta + BX\beta$.
- Because $\tilde{\beta}$ is unbiased, this must mean that $BX = 0$.
- The covariance matrix of $\tilde{\beta}$ is

Gauss-Markov Theorem and Proof

- We get back the least squares estimator if $A = (X^T X)^{-1} X^T$ so let us assume that $B := A - (X^T X)^{-1} X^T \neq 0$.
- Now the expectation of $\tilde{\beta}$ equals $\beta + BX\beta$.
- Because $\tilde{\beta}$ is unbiased, this must mean that $BX = 0$.
- The covariance matrix of $\tilde{\beta}$ is

$$Cov(\tilde{\beta}) = Cov(AY)$$

Gauss-Markov Theorem and Proof

- We get back the least squares estimator if $A = (X^T X)^{-1} X^T$ so let us assume that $B := A - (X^T X)^{-1} X^T \neq 0$.
- Now the expectation of $\tilde{\beta}$ equals $\beta + BX\beta$.
- Because $\tilde{\beta}$ is unbiased, this must mean that $BX = 0$.
- The covariance matrix of $\tilde{\beta}$ is

$$\begin{aligned} Cov(\tilde{\beta}) &= Cov(AY) \\ &= ACov(Y)A^T \end{aligned}$$

Gauss-Markov Theorem and Proof

- We get back the least squares estimator if $A = (X^T X)^{-1} X^T$ so let us assume that $B := A - (X^T X)^{-1} X^T \neq 0$.
- Now the expectation of $\tilde{\beta}$ equals $\beta + BX\beta$.
- Because $\tilde{\beta}$ is unbiased, this must mean that $BX = 0$.
- The covariance matrix of $\tilde{\beta}$ is

$$
\begin{aligned}
Cov(\tilde{\beta}) &= Cov(AY) \\
&= ACov(Y)A^T \\
&= [B + (X^T X)^{-1} X^T][\sigma^2 I_n][B + (X^T X)^{-1} X^T]^T
\end{aligned}
$$

Gauss-Markov Theorem and Proof

- We get back the least squares estimator if $A = (X^T X)^{-1} X^T$ so let us assume that $B := A - (X^T X)^{-1} X^T \neq 0$.
- Now the expectation of $\tilde{\beta}$ equals $\beta + BX\beta$.
- Because $\tilde{\beta}$ is unbiased, this must mean that $BX = 0$.
- The covariance matrix of $\tilde{\beta}$ is

$$\begin{aligned}
Cov(\tilde{\beta}) &= Cov(AY) \\
&= ACov(Y)A^T \\
&= [B + (X^T X)^{-1} X^T][\sigma^2 I_n][B + (X^T X)^{-1} X^T]^T \\
&= \sigma^2 [B + (X^T X)^{-1} X^T][B^T + X(X^T X)^{-1}]
\end{aligned}$$

Gauss-Markov Theorem and Proof

- We get back the least squares estimator if $A = (X^T X)^{-1} X^T$ so let us assume that $B := A - (X^T X)^{-1} X^T \neq 0$.
- Now the expectation of $\tilde{\beta}$ equals $\beta + BX\beta$.
- Because $\tilde{\beta}$ is unbiased, this must mean that $BX = 0$.
- The covariance matrix of $\tilde{\beta}$ is

$$\begin{aligned}
Cov(\tilde{\beta}) &= Cov(AY) \\
&= ACov(Y)A^T \\
&= [B + (X^T X)^{-1} X^T][\sigma^2 I_n][B + (X^T X)^{-1} X^T]^T \\
&= \sigma^2 [B + (X^T X)^{-1} X^T][B^T + X(X^T X)^{-1}] \\
&= \sigma^2 [(X^T X)^{-1} + BX(X^T X)^{-1} + (X^T X)^{-1} X^T B^T + \\
&\quad BB^T]
\end{aligned}$$

Gauss-Markov Theorem and Proof

- Because $BX = 0$, we deduce from the previous expression that
$$Cov(\tilde{\beta}) = \sigma^2(X^TX)^{-1} + \sigma^2BB^T.$$

Gauss-Markov Theorem and Proof

- Because $BX = 0$, we deduce from the previous expression that
$$Cov(\tilde{\beta}) = \sigma^2(X^TX)^{-1} + \sigma^2 BB^T.$$

- Recall that $\sigma^2(X^TX)^{-1}$ is the covariance matrix of the least squares estimator $\hat{\beta}$.

Gauss-Markov Theorem and Proof

- Because $BX = 0$, we deduce from the previous expression that
$$Cov(\tilde{\beta}) = \sigma^2 (X^T X)^{-1} + \sigma^2 BB^T.$$

- Recall that $\sigma^2 (X^T X)^{-1}$ is the covariance matrix of the least squares estimator $\hat{\beta}$. We thus have
$$Cov(\tilde{\beta}) = Cov(\hat{\beta}) + \sigma^2 BB^T.$$

Gauss-Markov Theorem and Proof

- Because $BX = 0$, we deduce from the previous expression that
$$Cov(\tilde{\beta}) = \sigma^2 (X^T X)^{-1} + \sigma^2 BB^T.$$

- Recall that $\sigma^2 (X^T X)^{-1}$ is the covariance matrix of the least squares estimator $\hat{\beta}$. We thus have
$$Cov(\tilde{\beta}) = Cov(\hat{\beta}) + \sigma^2 BB^T.$$

- It is important to note that the Gauss-Markov theorem requires the assumption of homoskedasticity.

The case when $X^T X$ is not invertible

The case when $X^T X$ is not invertible

- Note first that $X^T X$ is not invertible if and only if there exists a column of X that can be written as the linear combination of all the other columns.

The case when $X^T X$ is not invertible

- Note first that $X^T X$ is not invertible if and only if there exists a column of X that can be written as the linear combination of all the other columns.
- To see this, just note that $X^T X$ is non-invertible if and only if there exists a non-zero vector w for which $X^T X w = 0$ which is equivalent to $Xw = 0$.

The case when $X^T X$ is not invertible

- Note first that $X^T X$ is not invertible if and only if there exists a column of X that can be written as the linear combination of all the other columns.
- To see this, just note that $X^T X$ is non-invertible if and only if there exists a non-zero vector w for which $X^T X w = 0$ which is equivalent to $Xw = 0$.
- This is because

$$X^T X w = 0 \Rightarrow w^T X^T X w = 0 \Rightarrow \|Xw\|_2^2 = 0 \rightarrow Xw = 0.$$

The case when $X^T X$ is not invertible

- Note first that $X^T X$ is not invertible if and only if there exists a column of X that can be written as the linear combination of all the other columns.
- To see this, just note that $X^T X$ is non-invertible if and only if there exists a non-zero vector w for which $X^T X w = 0$ which is equivalent to $Xw = 0$.
- This is because

$$X^T X w = 0 \Rightarrow w^T X^T X w = 0 \Rightarrow \|Xw\|_2^2 = 0 \to Xw = 0.$$

- In the case when $X^T X$ is not invertible, we obviously cannot solve the normal equations $X^T X b = X^T Y$ as $b = (X^T X)^{-1} X^T Y$.

The case when $X^T X$ is not invertible

- Note first that $X^T X$ is not invertible if and only if there exists a column of X that can be written as the linear combination of all the other columns.
- To see this, just note that $X^T X$ is non-invertible if and only if there exists a non-zero vector w for which $X^T X w = 0$ which is equivalent to $Xw = 0$.
- This is because

$$X^T X w = 0 \Rightarrow w^T X^T X w = 0 \Rightarrow \|Xw\|_2^2 = 0 \rightarrow Xw = 0.$$

- In the case when $X^T X$ is not invertible, we obviously cannot solve the normal equations $X^T X b = X^T Y$ as $b = (X^T X)^{-1} X^T Y$.
- So how do we estimate β here?

The case when $X^T X$ is not invertible

- Note first that $X^T X$ is not invertible if and only if there exists a column of X that can be written as the linear combination of all the other columns.
- To see this, just note that $X^T X$ is non-invertible if and only if there exists a non-zero vector w for which $X^T X w = 0$ which is equivalent to $Xw = 0$.
- This is because

$$X^T X w = 0 \Rightarrow w^T X^T X w = 0 \Rightarrow \|Xw\|_2^2 = 0 \rightarrow Xw = 0.$$

- In the case when $X^T X$ is not invertible, we obviously cannot solve the normal equations $X^T X b = X^T Y$ as $b = (X^T X)^{-1} X^T Y$.
- So how do we estimate β here?
- It turns out that even when $X^T X$ is non-invertible, the normal equations have a solution.

The case when $X^T X$ is not invertible

- This is because the column space of $X^T X$ (denoted by $\mathcal{C}(X^T X)$) always equals the column space of X^T (denoted by $\mathcal{C}(X^T)$).

The case when $X^T X$ is not invertible

- This is because the column space of $X^T X$ (denoted by $\mathcal{C}(X^T X)$) always equals the column space of X^T (denoted by $\mathcal{C}(X^T)$).
- To see why $\mathcal{C}(X^T X) = \mathcal{C}(X^T)$, note that it is enough to prove that $\mathcal{C}(X^T X)^\perp = \mathcal{C}(X^T)^\perp$ where $\mathcal{C}(X^T X)^\perp$ denotes the orthogonal complement of $\mathcal{C}(X^T X)$ (similarly $\mathcal{C}(X^T)^\perp$ denotes the orthogonal complement of $\mathcal{C}(X^T)$).

The case when $X^T X$ is not invertible

- This is because the column space of $X^T X$ (denoted by $\mathcal{C}(X^T X)$) always equals the column space of X^T (denoted by $\mathcal{C}(X^T)$).
- To see why $\mathcal{C}(X^T X) = \mathcal{C}(X^T)$, note that it is enough to prove that $\mathcal{C}(X^T X)^\perp = \mathcal{C}(X^T)^\perp$ where $\mathcal{C}(X^T X)^\perp$ denotes the orthogonal complement of $\mathcal{C}(X^T X)$ (similarly $\mathcal{C}(X^T)^\perp$ denotes the orthogonal complement of $\mathcal{C}(X^T)$).
- Now a vector $u \in \mathcal{C}(X^T X)^\perp$ if and only if $u^T X^T X = 0$

The case when $X^T X$ is not invertible

- This is because the column space of $X^T X$ (denoted by $\mathcal{C}(X^T X)$) always equals the column space of X^T (denoted by $\mathcal{C}(X^T)$).
- To see why $\mathcal{C}(X^T X) = \mathcal{C}(X^T)$, note that it is enough to prove that $\mathcal{C}(X^T X)^\perp = \mathcal{C}(X^T)^\perp$ where $\mathcal{C}(X^T X)^\perp$ denotes the orthogonal complement of $\mathcal{C}(X^T X)$ (similarly $\mathcal{C}(X^T)^\perp$ denotes the orthogonal complement of $\mathcal{C}(X^T)$).
- Now a vector $u \in \mathcal{C}(X^T X)^\perp$ if and only if $u^T X^T X = 0$

The case when $X^T X$ is not invertible

- This is because the column space of $X^T X$ (denoted by $\mathcal{C}(X^T X)$) always equals the column space of X^T (denoted by $\mathcal{C}(X^T)$).
- To see why $\mathcal{C}(X^T X) = \mathcal{C}(X^T)$, note that it is enough to prove that $\mathcal{C}(X^T X)^\perp = \mathcal{C}(X^T)^\perp$ where $\mathcal{C}(X^T X)^\perp$ denotes the orthogonal complement of $\mathcal{C}(X^T X)$ (similarly $\mathcal{C}(X^T)^\perp$ denotes the orthogonal complement of $\mathcal{C}(X^T)$).
- Now a vector $u \in \mathcal{C}(X^T X)^\perp$ if and only if $u^T X^T X = 0$ which is the same as $u^T X^T = 0$ i.e., $u \in \mathcal{C}(X^T)$.

The case when $X^T X$ is not invertible

- This is because the column space of $X^T X$ (denoted by $\mathcal{C}(X^T X)$) always equals the column space of X^T (denoted by $\mathcal{C}(X^T)$).
- To see why $\mathcal{C}(X^T X) = \mathcal{C}(X^T)$, note that it is enough to prove that $\mathcal{C}(X^T X)^\perp = \mathcal{C}(X^T)^\perp$ where $\mathcal{C}(X^T X)^\perp$ denotes the orthogonal complement of $\mathcal{C}(X^T X)$ (similarly $\mathcal{C}(X^T)^\perp$ denotes the orthogonal complement of $\mathcal{C}(X^T)$).
- Now a vector $u \in \mathcal{C}(X^T X)^\perp$ if and only if $u^T X^T X = 0$ which is the same as $u^T X^T = 0$ i.e., $u \in \mathcal{C}(X^T)$.
- Because $\mathcal{C}(X^T X) = \mathcal{C}(X^T)$, the normal equations $X^T X b = X^T Y$ always have a solution

The case when $X^T X$ is not invertible

- This is because the column space of $X^T X$ (denoted by $\mathcal{C}(X^T X)$) always equals the column space of X^T (denoted by $\mathcal{C}(X^T)$).
- To see why $\mathcal{C}(X^T X) = \mathcal{C}(X^T)$, note that it is enough to prove that $\mathcal{C}(X^T X)^\perp = \mathcal{C}(X^T)^\perp$ where $\mathcal{C}(X^T X)^\perp$ denotes the orthogonal complement of $\mathcal{C}(X^T X)$ (similarly $\mathcal{C}(X^T)^\perp$ denotes the orthogonal complement of $\mathcal{C}(X^T)$).
- Now a vector $u \in \mathcal{C}(X^T X)^\perp$ if and only if $u^T X^T X = 0$ which is the same as $u^T X^T = 0$ i.e., $u \in \mathcal{C}(X^T)$.
- Because $\mathcal{C}(X^T X) = \mathcal{C}(X^T)$, the normal equations $X^T X b = X^T Y$ always have a solution (because $X^T Y \in \mathcal{C}(X^T) = \mathcal{C}(X^T X)$, we can always write $X^T Y$ as $X^T X b$ for some b).

The case when $X^T X$ is not invertible

- However, when $X^T X$ is non-invertible, the normal equations will have infinitely many solutions.

The case when $X^T X$ is not invertible

- However, when $X^T X$ is non-invertible, the normal equations will have infinitely many solutions. In this case, we shall call any solution of $X^T X b = X^T Y$ a least squares estimator of β.

The case when $X^T X$ is not invertible

- However, when $X^T X$ is non-invertible, the normal equations will have infinitely many solutions. In this case, we shall call any solution of $X^T X b = X^T Y$ a least squares estimator of β.
- When $X^T X$ is non-invertible, there will be infinitely many least squares estimators for β.

The case when $X^T X$ is not invertible

- However, when $X^T X$ is non-invertible, the normal equations will have infinitely many solutions. In this case, we shall call any solution of $X^T X b = X^T Y$ a least squares estimator of β.
- When $X^T X$ is non-invertible, there will be infinitely many least squares estimators for β.
- It might seem strange to have infinitely many least squares estimators. It turns out however that when $X^T X$ is non-invertible, estimation of β is not a well-defined problem.

The case when $X^T X$ is not invertible

- However, when $X^T X$ is non-invertible, the normal equations will have infinitely many solutions. In this case, we shall call any solution of $X^T X b = X^T Y$ a least squares estimator of β.
- When $X^T X$ is non-invertible, there will be infinitely many least squares estimators for β.
- It might seem strange to have infinitely many least squares estimators. It turns out however that when $X^T X$ is non-invertible, estimation of β is not a well-defined problem.
- To see this, note first that if X_{i*} denotes the ith row of X, then the ith observation Y_i can be written as $Y_i = X_{i*}\beta + e_i$. We therefore only have data on the linear combinations $X_{1*}\beta, \dots, X_{n*}\beta$.

The case when $X^T X$ is not invertible

- However, when $X^T X$ is non-invertible, the normal equations will have infinitely many solutions. In this case, we shall call any solution of $X^T X b = X^T Y$ a least squares estimator of β.
- When $X^T X$ is non-invertible, there will be infinitely many least squares estimators for β.
- It might seem strange to have infinitely many least squares estimators. It turns out however that when $X^T X$ is non-invertible, estimation of β is not a well-defined problem.
- To see this, note first that if X_{i*} denotes the ith row of X, then the ith observation Y_i can be written as $Y_i = X_{i*}\beta + e_i$. We therefore only have data on the linear combinations $X_{1*}\beta, \dots, X_{n*}\beta$.
- Therefore, we can only estimate a linear function $\lambda^T \beta$ if the vector λ can be written as a linear combination of the rows of X.

Estimability

- The previous discussion motivates the following definition of estimability:

Estimability

- The previous discussion motivates the following definition of estimability:

Definition

A linear function $\lambda^T\beta$ of β is said to be estimable if the vector λ is a linear combination of the rows of X.

Estimability

- The previous discussion motivates the following definition of estimability:

Definition

A linear function $\lambda^T\beta$ of β is said to be estimable if the vector λ is a linear combination of the rows of X.

- Equivalently, we say that $\lambda^T\beta$ is estimable if and only if $\lambda \in \mathcal{C}(X^T)$.

Estimability

- The previous discussion motivates the following definition of estimability:

Definition

A linear function $\lambda^T\beta$ of β is said to be estimable if the vector λ is a linear combination of the rows of X.

- Equivalently, we say that $\lambda^T\beta$ is estimable if and only if $\lambda \in \mathcal{C}(X^T)$.
- Because $\mathcal{C}(X^T) = \mathcal{C}(X^TX)$, estimability of $\lambda^T\beta$ is further equivalent to $\lambda \in \mathcal{C}(X^TX)$.

Estimability

- The previous discussion motivates the following definition of estimability:

Definition

A linear function $\lambda^T\beta$ of β is said to be estimable if the vector λ is a linear combination of the rows of X.

- Equivalently, we say that $\lambda^T\beta$ is estimable if and only if $\lambda \in \mathcal{C}(X^T)$.
- Because $\mathcal{C}(X^T) = \mathcal{C}(X^T X)$, estimability of $\lambda^T\beta$ is further equivalent to $\lambda \in \mathcal{C}(X^T X)$.
- It therefore only makes sense to estimate $\lambda\beta$ if $\lambda \in \mathcal{C}(X^T X)$.

Estimability

- The previous discussion motivates the following definition of estimability:

Definition

A linear function $\lambda^T\beta$ of β is said to be estimable if the vector λ is a linear combination of the rows of X.

- Equivalently, we say that $\lambda^T\beta$ is estimable if and only if $\lambda \in \mathcal{C}(X^T)$.
- Because $\mathcal{C}(X^T) = \mathcal{C}(X^TX)$, estimability of $\lambda^T\beta$ is further equivalent to $\lambda \in \mathcal{C}(X^TX)$.
- It therefore only makes sense to estimate $\lambda\beta$ if $\lambda \in \mathcal{C}(X^TX)$.
- If X^TX is invertible, then every $\lambda^T\beta$ is estimable.

Estimability

- The previous discussion motivates the following definition of estimability:

Definition

A linear function $\lambda^T \beta$ of β is said to be estimable if the vector λ is a linear combination of the rows of X.

- Equivalently, we say that $\lambda^T \beta$ is estimable if and only if $\lambda \in \mathcal{C}(X^T)$.
- Because $\mathcal{C}(X^T) = \mathcal{C}(X^T X)$, estimability of $\lambda^T \beta$ is further equivalent to $\lambda \in \mathcal{C}(X^T X)$.
- It therefore only makes sense to estimate $\lambda \beta$ if $\lambda \in \mathcal{C}(X^T X)$.
- If $X^T X$ is invertible, then every $\lambda^T \beta$ is estimable.
- If $X^T X$ is non-invertible, then not every $\lambda^T \beta$ is estimable.

Estimability

- **Result:** If $\lambda^T\beta$ is estimable, then $\lambda^T\hat{\beta}_{ls}$ is the same for every solution $\hat{\beta}_{ls}$ of the normal equations. In other words, the least squares estimate of $\lambda^T\beta$ is unique.

Estimability

- **Result:** If $\lambda^T \beta$ is estimable, then $\lambda^T \hat{\beta}_{ls}$ is the same for every solution $\hat{\beta}_{ls}$ of the normal equations. In other words, the least squares estimate of $\lambda^T \beta$ is unique.
- **Proof.** Since $\lambda^T \beta$ is estimable, the vector λ lies in the column space of $X^T X$ and hence $\lambda = X^T X u$ for some vector u.

Estimability

- **Result:** If $\lambda^T\beta$ is estimable, then $\lambda^T\hat{\beta}_{ls}$ is the same for every solution $\hat{\beta}_{ls}$ of the normal equations. In other words, the least squares estimate of $\lambda^T\beta$ is unique.
- **Proof.** Since $\lambda^T\beta$ is estimable, the vector λ lies in the column space of X^TX and hence $\lambda = X^TXu$ for some vector u. Therefore,

$$\lambda^T\hat{\beta}_{ls} = u^TX^TX\hat{\beta}_{ls}$$

Estimability

- **Result:** If $\lambda^T\beta$ is estimable, then $\lambda^T\hat{\beta}_{ls}$ is the same for every solution $\hat{\beta}_{ls}$ of the normal equations. In other words, the least squares estimate of $\lambda^T\beta$ is unique.
- **Proof.** Since $\lambda^T\beta$ is estimable, the vector λ lies in the column space of X^TX and hence $\lambda = X^TXu$ for some vector u. Therefore,

$$\lambda^T\hat{\beta}_{ls} = u^TX^TX\hat{\beta}_{ls} = u^TX^TY,$$

Estimability

- **Result:** If $\lambda^T \beta$ is estimable, then $\lambda^T \hat{\beta}_{ls}$ is the same for every solution $\hat{\beta}_{ls}$ of the normal equations. In other words, the least squares estimate of $\lambda^T \beta$ is unique.
- **Proof.** Since $\lambda^T \beta$ is estimable, the vector λ lies in the column space of $X^T X$ and hence $\lambda = X^T X u$ for some vector u. Therefore,

$$\lambda^T \hat{\beta}_{ls} = u^T X^T X \hat{\beta}_{ls} = u^T X^T Y,$$

where the last equality follows from the fact that $\hat{\beta}_{ls}$ satisfies the normal equations.

Estimability

- **Result:** If $\lambda^T \beta$ is estimable, then $\lambda^T \hat{\beta}_{ls}$ is the same for every solution $\hat{\beta}_{ls}$ of the normal equations. In other words, the least squares estimate of $\lambda^T \beta$ is unique.
- **Proof.** Since $\lambda^T \beta$ is estimable, the vector λ lies in the column space of $X^T X$ and hence $\lambda = X^T X u$ for some vector u. Therefore,

$$\lambda^T \hat{\beta}_{ls} = u^T X^T X \hat{\beta}_{ls} = u^T X^T Y,$$

where the last equality follows from the fact that $\hat{\beta}_{ls}$ satisfies the normal equations. Since u only depends on λ, this proves that $\lambda^T \hat{\beta}_{ls}$ does not depend on the particular choice of the solution $\hat{\beta}_{ls}$ of the normal equations.

Summary

The situation when $X^T X$ is non-invertible can be summarized as follows:

Summary

The situation when $X^T X$ is non-invertible can be summarized as follows:

- There are infinitely many solutions of the normal equations $X^T Xb = X^T Y$.

Summary

The situation when $X^T X$ is non-invertible can be summarized as follows:

- There are infinitely many solutions of the normal equations $X^T X b = X^T Y$.
- Not every linear combination $\lambda^T \beta$ can be estimated from regression data. We say that $\lambda^T \beta$ is estimable if and only if $\lambda \in \mathcal{C}(X^T) = \mathcal{C}(X^T X)$.

Summary

The situation when $X^T X$ is non-invertible can be summarized as follows:

- There are infinitely many solutions of the normal equations $X^T X b = X^T Y$.
- Not every linear combination $\lambda^T \beta$ can be estimated from regression data. We say that $\lambda^T \beta$ is estimable if and only if $\lambda \in \mathcal{C}(X^T) = \mathcal{C}(X^T X)$.
- When $\lambda^T \beta$ is estimable, the value $\lambda^T \hat{\beta}_{ls}$ is the same for every solution $\hat{\beta}_{ls}$ of the normal equations $X^T X b = X^T Y$.