OLS as Orthogonal Projection

date: 20241007
status: #finished
tags: #econometrics #economic_research #ols

---

1. OLS

(i) Closed-form solution

Definition (Ordinary Least Squares)

Given a full-column-rank matrix X and a vector y — arbitrary numbers, nothing assumed random — the ordinary least squares (OLS) solution is the coefficient vector that minimizes the squared residual:

$$\hat{\beta}^{\text{OLS}} = \arg\min_{\beta} \|y - X\beta\|^2 = (X'X)^{-1}X'y$$

Proof

$$\begin{aligned}
\frac{\partial}{\partial\beta}\|y - X\beta\|^2 &= \frac{\partial}{\partial\beta}(y - X\beta)'(y - X\beta) \\
&= \frac{\partial}{\partial\beta}(y'y - y'X\beta - \beta'X'y + \beta'X'X\beta) \\
&= -X'y - X'y + 2(X'X)\beta
\end{aligned}$$

FOC w.r.t β:

$$\left[\frac{\partial}{\partial\beta}\|y - X\beta\|^2\right]\Bigg|_{\beta=\beta^{\text{OLS}}} = 0$$

$$\Rightarrow \hat{\beta}^{\text{OLS}} = (X'X)^{-1}(X'y)$$

(ii) Existence

Theorem

If column vectors of X are linearly independent, i.e., X has full column rank, then

1. *$\hat{\beta}^{\text{OLS}}$ is a minimizer*

2. *the inverse of $X'X$ exists; therefore, $\hat{\beta}^{\text{OLS}}$ exists.*

Proof (Proof of $\hat{\beta}^{\text{OLS}}$ is a minimizer)

SOC:

$$2X'X > 0$$

which means $X'X$ should be positive definite for $\hat{\beta}^{\text{OLS}}$ to be a valid minimizer.

Lemma

$X'X$ is naturally semi-positive definite. If column vectors are linearly independent, it's positive definite.

Proof

For any non-zero vector v,

$$v'(X'X)v = (Xv)'Xv = \|Xv\|^2 \geq 0$$

Therefore, $X'X$ is naturally semi-definite.

Given that v is non-zero, and by the definition of linear independent, i.e., $Xv \neq 0$ for all $v \neq 0$, Therefore, $\|Xv\|^2 > 0$, implying $X'X$ is positive definite

By the above lemma, if X is linearly independent, then $X'X$ is positive definite, and then SOC is satisfied, thereby making $\hat{\beta}^{\text{OLS}}$ a minimizer.

Proof (Proof of the existence of $(X'X)^{-1}$)

Lemma

If columns vectors of X are linearly independent, then column vectors of $X'X$ are also linearly independent

Proof

Suppose $(X'X)v = 0$. Then, $v'(X'X)v = 0$.

$$v'(X'X)v = (Xv)'(Xv) = \|Xv\|^2 = 0,$$

implying $Xv = 0$. As column vectors of X are linearly independent, $Xv = 0$ implies $v = 0$.

Consequently, column vectors of $X'X$ are linearly independent.

Therefore, by the above lemma, if column vectors of X are linearly independent, then $(X'X)^{-1}$ exist as column vectors are linearly independent and we can obtain $\hat{\beta}^{\text{OLS}} = (X'X)^{-1}X'y$.

(iii) Uniqueness: β is identified

Full column rank of X not only guarantees the existence of $\hat{\beta}^{\text{OLS}}$ but also ensures its uniqueness.

Definition (Identification)

The coefficient vector $\beta \in \mathbb{R}^k$ is **identified** when distinct coefficient vectors produce distinct fits — that is, for all $\beta, \gamma \in \mathbb{R}^k$,

$$X\gamma = X\beta \Rightarrow \gamma = \beta$$

Theorem

X has full column rank if and only if β is identified.

Proof (Sufficiency: full column rank implies identification)

Suppose X has full column rank and $X\beta = X\gamma$. Then

$$X(\beta - \gamma) = 0$$

As X has full column rank, $\beta - \gamma = 0$, i.e., $\beta = \gamma$. Therefore, β is identified.

Proof (Necessity: identification implies full column rank)

If column vectors of X are **not** linearly independent, then there exists a column X_i and $\alpha \in \mathbb{R}^{k-1}$ such that $X_i = X_{-i}\alpha$, where X_{-i} denotes the matrix excluding column X_i. Consequently, we can find another parameter vector

$$\gamma = \begin{pmatrix} \beta_1 + \alpha_1 \beta_i \\ \beta_2 + \alpha_2 \beta_i \\ \vdots \\ \beta_{i-1} + \alpha_{i-1} \beta_i \\ 0 \\ \beta_{i+1} + \alpha_{i+1} \beta_i \\ \vdots \\ \beta_k + \alpha_k \beta_i \end{pmatrix} \in \mathbb{R}^k$$

such that $X\beta = X\gamma$. Therefore, we have proved that X **not** having full column rank implies β is unidentified. Equivalently, β being identified implies X has full column rank.

(iv) Huge Dataset

On huge datasets where the row size is large the closed form is computed by streaming the sufficient statistics $X'X$ and $X'y$ chunk by chunk — the full X is never needed in memory.

---

2. Properties of OLS

(i) Moment condition

Theorem (OLS's Moment Condition)

FOC implies that

$$X'e = 0$$

where $e = y - X\beta^{\mathrm{OLS}}$.

Proof

FOC states that $X'y = X'X\beta^{\mathrm{OLS}}$ and

$$X'y = X'\left(X\beta^{\mathrm{OLS}} + e\right) = X'X\beta^{\mathrm{OLS}} + X'e$$

$\hat{\beta}^{\mathrm{OLS}} = (X'X)^{-1}X'y$ implies that $X\hat{\beta}^{\mathrm{OLS}}$ is the projection of y on the column space of X. If y is the projection, than $e = y - X\hat{\beta}$ should be orthogonal to all column vectors of X $(X_1, X_2, \ldots, X_N)$, which is implied by:

$$X'e = \begin{pmatrix} X_1'e \\ X_2'e \\ \dots \\ X_N'e \end{pmatrix} = 0$$

(ii) Residuals sum to zero

Corollary

Suppose $\mathbb{1}_n \in \mathbb{R}^n$ is one of the columns of X,

$$\sum_{i=1}^{N} e_i = 0$$

Proof

Since $\mathbb{1}_n$ is a column of X, the moment condition gives $\mathbb{1}_n' e = \sum_{i=1}^{N} e_i = 0$.

(iii) The fit passes through the sample means

Theorem (OLS pass the mean $(\bar{y}, \bar{X})$)

$$\bar{y} = \bar{X}\beta^{\text{OLS}}$$

Proof

$$\begin{aligned} y &= X\beta^{\text{OLS}} + e \\ \frac{1}{n}\mathbb{1}'y &= \frac{1}{n}\mathbb{1}'X\beta^{\text{OLS}} + \frac{1}{n}\mathbb{1}'e \\ \bar{y} &= \bar{X}\beta^{\text{OLS}} + \frac{1}{n}\sum_{i=1}^{N} e_i \end{aligned}$$

As $\sum_{i=1}^{N} e_i = 0$ is proved above, $\bar{y} = \bar{X}\beta^{\text{OLS}}$.