Linear Regression Basics

Chungsang Tom Lam[1]

[1]Department of Economics
Clemson University

January 30, 2019

The Model

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & \dots & x_{1(k-1)} \\ 1 & x_{21} & x_{22} & x_{23} & \dots & x_{2(k-1)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} & \dots & x_{n(k-1)} \end{bmatrix}_{n\times k} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{k-1} \end{bmatrix}_{k\times 1} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \vdots \\ \epsilon_n \end{bmatrix}_{n\times 1}$$

$$y = X\beta + \epsilon$$

The Estimators

- The estimator of β is $\hat{\beta}$
- The residual vector is $\hat{\epsilon} = y - X\hat{\beta}$
- The sum of squared residuals (SSR) is $\sum_{i=1}^{n} \hat{\epsilon}_i^2$ or $\hat{\epsilon}'\hat{\epsilon}$

$$\begin{bmatrix} \hat{\epsilon}_1 & \hat{\epsilon}_2 & \ldots & \hat{\epsilon}_n \end{bmatrix}_{1\times n} \begin{bmatrix} \hat{\epsilon}_1 \\ \hat{\epsilon}_2 \\ \vdots \\ \vdots \\ \hat{\epsilon}_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} \hat{\epsilon}_1\hat{\epsilon}_1 + \hat{\epsilon}_2\hat{\epsilon}_2 + \cdots + \hat{\epsilon}_n\hat{\epsilon}_n \end{bmatrix}_{1\times 1}$$

- The fitted value is $\hat{y} = X\hat{\beta}$

Sum of squared residuals

$$
\begin{aligned}
\hat{\epsilon}'\hat{\epsilon} &= (y - X\hat{\beta})'(y - X\hat{\beta}) \\
&= y'y - \hat{\beta}'X'y - y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta} \\
&= y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}
\end{aligned}
$$

To minimize the sum of squared residuals, differentiate w.r.t. $\hat{\beta}$

$$
\begin{aligned}
\frac{\partial \hat{\epsilon}'\hat{\epsilon}}{\partial \hat{\beta}} &= \frac{\partial}{\partial \hat{\beta}}(y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}) \\
&= -2X'y + 2X'X\hat{\beta} = 0
\end{aligned}
$$

Solution of $\hat{\beta}$

$$
\begin{aligned}
-2X'y + 2X'X\hat{\beta} &= 0 \\
X'X\hat{\beta} &= X'y \\
(X'X)^{-1}X'X\hat{\beta} &= (X'X)^{-1}X'y \\
\hat{\beta} &= (X'X)^{-1}X'y
\end{aligned}
$$

OLS assumptions

- $y = X\beta + \epsilon$
- X is full rank in another word, no strict multicollinearity an explanatory variable cannot have linear relationship of another explanatory variable.
- $E[\epsilon|X] = 0$
- $E[\epsilon\epsilon'|X] = \sigma^2 I$

$$\hat{\sigma}^2 = \frac{e'e}{\underbrace{N-K}_{\text{degrees of freedom}}}$$

$$\begin{bmatrix} \sigma^2 & & & & \\ & \sigma^2 & & & 0 \\ & & \sigma^2 & & \\ 0 & & & \sigma^2 & \\ & & & & \sigma^2 \end{bmatrix}$$

Unbiasness of OLS estimator

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y \\ &= (X'X)^{-1}X'(X\beta + \epsilon) \\ &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\epsilon \\ &= \beta + (X'X)^{-1}X'\epsilon\end{aligned}$$

$$\begin{aligned}E[\hat{\beta}|X] &= E[\beta + (X'X)^{-1}X'\epsilon|X] \\ &= \beta + E[(X'X)^{-1}X'\epsilon|X] \\ &= \beta\end{aligned}$$

Consistency of OLS estimator

$$
\begin{aligned}
plim_{n\to\infty}\hat{\beta} &= plim_{n\to\infty}\beta + (X'X)^{-1}X'\epsilon \\
&= \beta + plim_{n\to\infty}(X'X)^{-1}X'\epsilon \\
&= \beta + plim_{n\to\infty}(X'X)^{-1}plim_{n\to\infty}X'\epsilon \\
&= \beta + plim_{n\to\infty}(\frac{1}{n}X'X)^{-1}plim_{n\to\infty}(\frac{1}{n}X'\epsilon) \\
&= \beta
\end{aligned}
$$

Variance of OLS estimator

$$
\begin{aligned}
E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'|X] &= E[((X'X)^{-1}X'\epsilon)((X'X)^{-1}X'\epsilon)'|X] \\
&= E[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}] \\
&= (X'X)^{-1}X'E[\epsilon\epsilon'|X]X(X'X)^{-1} \\
&= (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1} \\
&= \sigma^2 I(X'X)^{-1}X'X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}
\end{aligned}
$$

$(\hat{\beta}-\beta)(\hat{\beta}-\beta)'$ is a k*k matrix

$\epsilon\epsilon'$ is a n*n matrix

Gauss Markov Theorem

Suppose there is an alternative unbiased linear estimator $\tilde{\beta} = Ay$ where

$$A = (X'X)^{-1}X' + D$$

$$\begin{aligned}
E[\tilde{\beta}|X] &= E[Ay|X] \\
&= E[((X'X)^{-1}X' + D)(X\beta + \epsilon)|X] \\
&= E[((X'X)^{-1}X' + D)X\beta + ((X'X)^{-1}X' + D)\epsilon)|X] \\
&= ((X'X)^{-1}X' + D)X\beta + E[((X'X)^{-1}X' + D)\epsilon)|X] \\
&= ((X'X)^{-1}X' + D)X\beta + E[(X'X)^{-1}X'\epsilon|X] + DE[\epsilon|X] \\
&= (I + DX)\beta
\end{aligned}$$

Which implies $DX = 0$ because $\tilde{\beta}$ is unbiased.

Gauss Markov Theorem

$$
\begin{aligned}
Var(\tilde{\beta}) =& Var(Ay) = AVar(y)A' \\
=& \sigma^2 AA' \\
=& \sigma^2((X'X)^{-1}X' + D)((X'X)^{-1}X' + D)' \\
=& \sigma^2((X'X)^{-1}X'(X'X)^{-1}X' + D(X'X)^{-1}X' \\
& + (X'X)^{-1}X'D' + DD') \\
=& \sigma^2((X'X)^{-1} + (X'X)^{-1}(DX)' + DX(X'X)^{-1} + DD') \\
=& \sigma^2((X'X)^{-1} + DD') \\
=& \sigma^2(X'X)^{-1} + \sigma^2 DD' \\
>& \sigma^2(X'X)^{-1}
\end{aligned}
$$

Estimator for $Var(\hat{\beta})$

Recall

$$E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'|X] = \sigma^2(X'X)^{-1}$$

We estimate σ^2 by $s^2 = \frac{\hat{\epsilon}'\hat{\epsilon}}{n-k}$ And the estimated variances of the elements in β are the diagonal elements of $s^2(X'X)^{-1}$
And hence we can construct the t-statistics for β

$$t = \frac{|\hat{\beta}-\beta_0|}{se(\hat{\beta})}$$