Econometric Analysis of Cross Section and Panel Data

Lecture 2: The Algebra of Least Squares

Zhian Hu

Central University of Finance and Economics

Fall 2024

This Lecture

- Hansen (2022): Chapter 3
- We introduce the popular least squares estimator.
- Most of the discussion will be algebraic, with questions of distribution and inference deferred to later lectures.

Samples

- The sample or dataset is $\{(Y_i, X_i) : i = 1, \ldots, n\}$, which are realizations of random variables $(Y, X) \in \mathbb{R} \times \mathbb{R}^k$.
  - Each (Y_i, X_i) is an observation.
  - A dataset is typically organized as a table where each column is a variable and each row is an observation.
- Assume $\{(Y_1, X_1), \ldots, (Y_i, X_i), \ldots, (Y_n, X_n)\}$ are independently and identically distributed.

Least Squares Estimator

- Consider the linear regression

error, not residual

$$Y_i = X_i'\beta + e_i$$

- The object of the least square estimator is to minimize

$$\widehat{S}(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - X_i'\beta\right)^2 = \frac{1}{n}\mathrm{SSE}(\beta)$$

  where $\mathrm{SSE}(\beta) = \sum_{i=1}^{n}\left(Y_i - X_i'\beta\right)^2$ is called the sum of squared errors function.
- We define the **least squares estimator** $\widehat{\beta}$ as the minimizer of $\widehat{S}(\beta)$. It is commonly called the **ordinary least squares (OLS)** estimator.
- **Important! The sample estimator $\widehat{S}(\beta)$ is a random feature of a random sample.**

Solving for Least Squares with One Regressor

最小化残差平方和

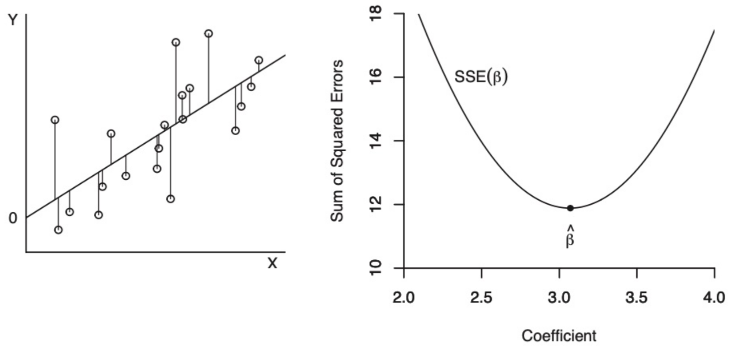
- Consider the case where there is only a **scalar** regressor X and a **scalar** coefficient β.

$$\mathrm{SSE}(\beta) = \sum_{i=1}^{n} (Y_i - X_i\beta)^2 = \left(\sum_{i=1}^{n} Y_i^2\right) - 2\beta\left(\sum_{i=1}^{n} X_i Y_i\right) + \beta^2\left(\sum_{i=1}^{n} X_i^2\right)$$

- The OLS estimator $\widehat{\beta}$ minimizes this function. Thus we have

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$$

Solving for Least Squares with One Regressor



(a) Deviation from Fitted Line

(b) Sum of Squared Error Function

Figure 3.1: Regression With One Regressor

Solving for Least Squares with Multiple Regressors

- Consider the case with multiple regressors and $\beta \in \mathbb{R}^k$ is a vector.

$$\text{SSE}(\beta) = \sum_{i=1}^{n} Y_i^2 - 2\beta' \sum_{i=1}^{n} X_i Y_i + \beta' \sum_{i=1}^{n} X_i X_i' \beta$$

- The OLS estimator $\widehat{\beta}$ minimizes this function. Thus we have

Q_{xx}=E[XX^T]总体矩

$$\sum_{i=1}^{n} X_i X_i' \widehat{\beta} = \sum_{i=1}^{n} X_i Y_i$$

矩估计：
黑箱p=\frac{1}{n}\sum A_i \
approx E[A](一阶原点矩)

$$\widehat{\beta} = \left(\sum_{i=1}^{n} X_i X_i' \right)^{-1} \left(\sum_{i=1}^{n} X_i Y_i \right) = \widehat{\boldsymbol{Q}}_{XX}^{-1} \widehat{\boldsymbol{Q}}_{XY}$$

if we define $\widehat{\boldsymbol{Q}}_{XY} = \frac{1}{n} \sum_{i=1}^{n} X_i Y_i$ and $\widehat{\boldsymbol{Q}}_{XX} = \frac{1}{n} \sum_{i=1}^{n} X_i X_i'$

样本距

Least Squares Residuals

- We define the fitted value $\widehat{Y}_i = X_i'\widehat{\beta}$ and the residual

$$\widehat{e}_i = Y_i - \widehat{Y}_i = Y_i - X_i'\widehat{\beta}$$

e_i is \beta without hat

- The **error** e_i is unobservable while the **residual** $\widehat{e}_i$ is an estimator.

$$\begin{aligned}\sum_{i=1}^{n} X_i \widehat{e}_i &= \sum_{i=1}^{n} X_i \left(Y_i - X_i'\widehat{\beta}\right) = \sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i X_i'\widehat{\beta} \\ &= \sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i X_i' \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} \left(\sum_{i=1}^{n} X_i Y_i\right) \\ &= \sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i Y_i = 0\end{aligned}$$

vector

- **When X_i contains a constant**, then $\frac{1}{n}\sum_{i=1}^{n} \widehat{e}_i = 0$.

Demeaned Regressors

- Write the linear regression with a constant as $Y_i = X_i'\beta + \alpha + e_i$.
- Minimize $SSE(\beta)$ and obtain the first order condition:

$$\sum_{i=1}^{n}\left(Y_i - X_i'\widehat{\beta} - \widehat{\alpha}\right) = 0, \sum_{i=1}^{n} X_i\left(Y_i - X_i'\widehat{\beta} - \widehat{\alpha}\right) \quad = 0$$

derivate to \alpha

derivate to \beta

- The first equation implies $\widehat{\alpha} = \bar{Y} - \bar{X}'\widehat{\beta}$

Demeaned Regressors

- Subtracting from the second we obtain

$$\sum_{i=1}^{n} X_i \left(\left(Y_i - \bar{Y}\right) - \left(X_i - \bar{X}\right)' \widehat{\beta} \right) = 0$$

- Then we find

$$\begin{aligned} \widehat{\beta} &= \left(\sum_{i=1}^{n} X_i \left(X_i - \bar{X}\right)' \right)^{-1} \left(\sum_{i=1}^{n} X_i \left(Y_i - \bar{Y}\right) \right) \\ &= \left(\sum_{i=1}^{n} \left(X_i - \bar{X}\right) \left(X_i - \bar{X}\right)' \right)^{-1} \left(\sum_{i=1}^{n} \left(X_i - \bar{X}\right) \left(Y_i - \bar{Y}\right) \right) \end{aligned}$$

homework

Model in Matrix Notation

- The n linear equations $Y_i = X_i'\beta + e_i$ make a system of n equations.

$$\begin{aligned} Y_1 &= X_1'\beta + e_1 \\ Y_2 &= X_2'\beta + e_2 \\ &\vdots \\ Y_n &= X_n'\beta + e_n \end{aligned}$$

- Define

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} X_1' \\ X_2' \\ \vdots \\ X_n' \end{pmatrix}, \quad \boldsymbol{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

- $\boldsymbol{Y}$ and $\boldsymbol{e}$ are $n \times 1$ vectors and $\boldsymbol{X}$ is an $n \times k$ matrix.

Model in Matrix Notation

- The system of n equations can be compactly written in the single equation

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{e}$$

- Sample sums can be written in matrix notation.

$$\sum_{i=1}^{n} X_i X_i' = \boldsymbol{X}'\boldsymbol{X}, \sum_{i=1}^{n} X_i Y_i = \boldsymbol{X}'\boldsymbol{Y}$$

- Therefore the least squares estimator can be written as

$$\widehat{\beta} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{Y}\right)$$

- Then we have $\boldsymbol{Y} = \boldsymbol{X}\widehat{\beta} + \widehat{\boldsymbol{e}}$ and $\boldsymbol{X}'\hat{\boldsymbol{e}} = 0$.
- The sum of squared error criterion is $\mathrm{SSE}(\beta) = (\boldsymbol{Y} - \boldsymbol{X}\beta)'(\boldsymbol{Y} - \boldsymbol{X}\beta)$

Projection Matrix

- Projection matrix: $\boldsymbol{P} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'$
- $\boldsymbol{P}\boldsymbol{Y} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y} = \boldsymbol{X}\widehat{\beta} = \widehat{\boldsymbol{Y}}$
- Properties of the projection matrix:
  1. $\boldsymbol{P}\boldsymbol{X} = \boldsymbol{X}$
  2. $\boldsymbol{P}$ is symmetric ($\boldsymbol{P}' = \boldsymbol{P}$).
  3. $\boldsymbol{P}$ is idempotent ($\boldsymbol{P}\boldsymbol{P} = \boldsymbol{P}$).
  4. $\operatorname{tr}\boldsymbol{P} = k$.
  5. The eigenvalues of $\boldsymbol{P}$ are 1 and 0 .
  6. $\boldsymbol{P}$ has k eigenvalues equalling 1 and $n-k$ equalling 0 .
  7. $\operatorname{rank}(\boldsymbol{P}) = k$
- Consider an example $\boldsymbol{P} = \mathbf{1}_n\left(\mathbf{1}_n'\mathbf{1}_n\right)^{-1}\mathbf{1}_n' = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$

$$\boldsymbol{P}\boldsymbol{Y} = \mathbf{1}_n\left(\mathbf{1}_n'\mathbf{1}_n\right)^{-1}\mathbf{1}_n'\boldsymbol{Y} = \mathbf{1}_n\bar{Y}$$

Annihilator Matrix

- Annihilator Matrix: $\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{P} = \boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'$
- $\boldsymbol{MY} = \boldsymbol{Y} - \boldsymbol{PY} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\beta} = \widehat{\boldsymbol{e}}$
- Properties of the annihilator matrix:

X完全拟合自身

  1. $\boldsymbol{MX} = 0$
  2. $\boldsymbol{M}$ is symmetric ($\boldsymbol{M}' = \boldsymbol{M}$).
  3. $\boldsymbol{M}$ is idempotent ($\boldsymbol{MM} = \boldsymbol{M}$).
  4. $\operatorname{tr}\boldsymbol{M} = n - k$.
- Consider the case $\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{P} = \boldsymbol{I}_n - \mathbf{1}_n\left(\mathbf{1}_n'\mathbf{1}_n\right)^{-1}\mathbf{1}_n'$, we have the demeaned value of Y_i.

$$\boldsymbol{MY} = \boldsymbol{Y} - \mathbf{1}_n\bar{Y}$$

去均值

- An alternative expression for the residual vector: $\widehat{\boldsymbol{e}} = \boldsymbol{MY} = \boldsymbol{M}(\boldsymbol{X}\beta + \boldsymbol{e}) = \boldsymbol{Me}$

误差和残差的关系

Estimation of Error Variance

- Besides β, another parameter we are interested in is the error variance $\sigma^2 = \mathbb{E}[e^2]$.
- A natural estimator for σ^2 is $\widetilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2$. 和P值有关
  - However, this is infeasible as e_i is not observed.
- The feasible estimator is $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \widehat{e}_i^2$.
- In matrix notation, $\widetilde{\sigma}^2 = n^{-1}\boldsymbol{e}'\boldsymbol{e}$ and

$$\widehat{\sigma}^2 = n^{-1}\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}} = n^{-1}\boldsymbol{e}'\boldsymbol{M}\boldsymbol{M}\boldsymbol{e} = n^{-1}\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e}$$

(Me)^{-1}(Me)

- An interesting implication is that

P对称幂等

$$\widetilde{\sigma}^2 - \widehat{\sigma}^2 = n^{-1}\boldsymbol{e}'\boldsymbol{e} - n^{-1}\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e} = n^{-1}\boldsymbol{e}'\boldsymbol{P}\boldsymbol{e} \geq 0$$

=eP^TPe=(Pe)^T(Pe)

e^T[I_n-M]e

- This shows that the feasible estimator is numerically smaller than the idealized estimator.

Analysis of Variance

- Use projection and annihilator matrix, we can write $\boldsymbol{Y} = \boldsymbol{PY} + \boldsymbol{MY} = \widehat{\boldsymbol{Y}} + \widehat{\boldsymbol{e}}$
- This decomposition is orthogonal, that is

$$\widehat{\boldsymbol{Y}}'\widehat{\boldsymbol{e}} = (\boldsymbol{PY})'(\boldsymbol{MY}) = \boldsymbol{Y}'\boldsymbol{PMY} = 0$$

PM=0

- Subtracting $\bar{Y}$ from both sides, we obtain

$$\boldsymbol{Y} - \mathbf{1}_n\bar{Y} = \widehat{\boldsymbol{Y}} - \mathbf{1}_n\bar{Y} + \widehat{\boldsymbol{e}}$$

Analysis of Variance

- This decomposition is also orthogonal when $\boldsymbol{X}$ contains a constant, as

$$\left(\widehat{\boldsymbol{Y}} - \mathbf{1}_n \bar{Y}\right)' \widehat{\boldsymbol{e}} = \widehat{\boldsymbol{Y}}' \widehat{\boldsymbol{e}} - \bar{Y} \mathbf{1}_n' \widehat{\boldsymbol{e}} = 0$$

前提是X包含常数项,此时有\sum e_i=0,即1_n\hat{e_i}

- It follows that

R^2

$$\left(\boldsymbol{Y} - \mathbf{1}_n \bar{Y}\right)' \left(\boldsymbol{Y} - \mathbf{1}_n \bar{Y}\right) = \left(\widehat{\boldsymbol{Y}} - \mathbf{1}_n \bar{Y}\right)' \left(\widehat{\boldsymbol{Y}} - \mathbf{1}_n \bar{Y}\right) + \widehat{\boldsymbol{e}}' \widehat{\boldsymbol{e}}$$

SST

or

$$\sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2 = \sum_{i=1}^{n} \left(\widehat{Y}_i - \bar{Y}\right)^2 + \sum_{i=1}^{n} \hat{e}_i^2$$

SSR

- This is commonly called the analysis-of-variance formula for least squares regression.

SSE?(\hat{Y_i}-\overline{\hat{Y_i}})，这两个square相等，因为Y=\hat{Y}+e_i,而\overline{Y}=\sum Y，逐步展开即可

Analysis of Variance

- A commonly reported statistic is the coefficient of determination or **R-squared**:

$$R^2 = \frac{\sum_{i=1}^n \left(\widehat{Y}_i - \bar{Y}\right)^2}{\sum_{i=1}^n \left(Y_i - \bar{Y}\right)^2} = 1 - \frac{\sum_{i=1}^n \widehat{e}_i^2}{\sum_{i=1}^n \left(Y_i - \bar{Y}\right)^2}$$

- It is often described as "the fraction of the sample variance of Y which is explained by the least squares fit".
- One deficiency with R^2 is that it increases when regressors are added to a regression so the "fit" can be always increased by increasing the number of regressors.

为何随着解释变量的增加而R^2不减？总平方和不变而残差平方和变小。\beta_n=0作为一种约束，体现变量的增减

Regression Components

- Partition $X = [X_1 \quad X_2]$ and $\beta = (\beta_1, \beta_2)$. The regression model can be written as

$$\boldsymbol{Y} = \boldsymbol{X}_1\beta_1 + \boldsymbol{X}_2\beta_2 + \boldsymbol{e}$$

- So that

$$\left(\widehat{\beta}_1, \widehat{\beta}_2\right) = \underset{\beta_1, \beta_2}{\operatorname{argmin}} \operatorname{SSE}(\beta_1, \beta_2)$$

  where

$$\operatorname{SSE}(\beta_1, \beta_2) = (\boldsymbol{Y} - \boldsymbol{X}_1\beta_1 - \boldsymbol{X}_2\beta_2)'(\boldsymbol{Y} - \boldsymbol{X}_1\beta_1 - \boldsymbol{X}_2\beta_2)$$

- Let's first focus on β_1. The solution can be written as

$$\widehat{\beta}_1 = \underset{\beta_1}{\operatorname{argmin}} \left(\min_{\beta_2} \operatorname{SSE}(\beta_1, \beta_2) \right)$$

记\overline{Y}=Y-X_i\beta_1
计算得出\beta_2回代得出残差平方和

- The inner expression $\min_{\beta_2} \operatorname{SSE}(\beta_1, \beta_2)$ minimizes over β_2 while holding β_1 fixed.

Residual Components

- Examine the inner minimization problem. This is simply the least squares regression of $Y - X_1\beta_1$ on X_2.This has solution

$$\underset{\beta_2}{\operatorname{argmin}}\, \mathrm{SSE}\left(\beta_1, \beta_2\right) = \left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\left(\boldsymbol{X}_2'\left(\boldsymbol{Y} - \boldsymbol{X}_1\beta_1\right)\right)$$

- Residuals equal to $\boldsymbol{M}_2\left(\boldsymbol{Y} - \boldsymbol{X}_1\beta_1\right)$, where $\boldsymbol{M}_2 = \boldsymbol{I}_n - \boldsymbol{X}_2\left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'$
- So the minimized value is

$$\begin{aligned}\min_{\beta_2} \mathrm{SSE}\left(\beta_1, \beta_2\right) &= \left(\boldsymbol{Y} - \boldsymbol{X}_1\beta_1\right)'\boldsymbol{M}_2\boldsymbol{M}_2\left(\boldsymbol{Y} - \boldsymbol{X}_1\beta_1\right) \\ &= \left(\boldsymbol{Y} - \boldsymbol{X}_1\beta_1\right)'\boldsymbol{M}_2\left(\boldsymbol{Y} - \boldsymbol{X}_1\beta_1\right)\end{aligned}$$

一直在简化目标函数

Residual Components

- Then we have

$$\widehat{\beta}_1 = \underset{\beta_1}{\operatorname{argmin}} \left(\boldsymbol{Y} - \boldsymbol{X}_1 \beta_1 \right)' \boldsymbol{M}_2 \left(\boldsymbol{Y} - \boldsymbol{X}_1 \beta_1 \right)$$
$$= \left(\boldsymbol{X}_1' \boldsymbol{M}_2 \boldsymbol{X}_1 \right)^{-1} \left(\boldsymbol{X}_1' \boldsymbol{M}_2 \boldsymbol{Y} \right)$$

- By a similar argument we find 不建议回代

$$\widehat{\beta}_2 = \left(\boldsymbol{X}_2' \boldsymbol{M}_1 \boldsymbol{X}_2 \right)^{-1} \left(\boldsymbol{X}_2' \boldsymbol{M}_1 \boldsymbol{Y} \right)$$

where $\boldsymbol{M}_1 = \boldsymbol{I}_n - \boldsymbol{X}_1 \left(\boldsymbol{X}_1' \boldsymbol{X}_1 \right)^{-1} \boldsymbol{X}_1'$.

Residual Regression

- Taken from the previous slide,

$$
\begin{aligned}
\widehat{\beta}_2 &= \left(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2\right)^{-1}\left(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{Y}\right) \\
&= \left(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{M}_1\boldsymbol{X}_2\right)^{-1}\left(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{M}_1\boldsymbol{Y}\right) \\
&= \left(\widetilde{\boldsymbol{X}}_2'\widetilde{\boldsymbol{X}}_2\right)^{-1}\left(\widetilde{\boldsymbol{X}}_2'\widetilde{\boldsymbol{e}}_1\right)
\end{aligned}
$$

两个残差项相乘，即OLS估计系数

where $\widetilde{\boldsymbol{X}}_2 = \boldsymbol{M}_1\boldsymbol{X}_2$ and $\widetilde{\boldsymbol{e}}_1 = \boldsymbol{M}_1\boldsymbol{Y}$. 正交矩阵乘Y即Y对x_1作OLS估计的残差项

- **Frisch-Waugh-Lovell (FWL) Theorem.** In $\boldsymbol{Y} = \boldsymbol{X}_1\beta_1 + \boldsymbol{X}_2\beta_2 + \boldsymbol{e}$, the OLS estimator of β_2 and the OLS residuals $\widehat{\boldsymbol{e}}$ may be computed via the following algorithm:
  1. Regress $\boldsymbol{Y}$ on $\boldsymbol{X}_1$, obtain residuals $\widetilde{\boldsymbol{e}}_1$;
  2. Regress $\boldsymbol{X}_2$ on $\boldsymbol{X}_1$, obtain residuals $\widetilde{\boldsymbol{X}}_2$; k个变量k个回归，残差项矩阵
  3. Regress $\widetilde{\boldsymbol{e}}_1$ on $\widetilde{\boldsymbol{X}}_2$, obtain OLS estimates $\widehat{\beta}_2$ and residuals $\widehat{\boldsymbol{e}}$.
- Can we obtain same results in two step? 双重差分中使用

Leverage Values 判断异常值

- The leverage values for the regressor matrix X are the diagonal elements of the projection matrix $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. Since

$$\boldsymbol{P} = \begin{pmatrix} X_1' \\ X_2' \\ \vdots \\ X_n' \end{pmatrix} (\boldsymbol{X}'\boldsymbol{X})^{-1} \begin{pmatrix} X_1 & X_2 & \cdots & X_n \end{pmatrix}$$

投影矩阵

- The leverage values are $h_{ii} = X_i'(\boldsymbol{X}'\boldsymbol{X})^{-1}X_i$. It measures how unusual the i^{th} observation X_i is relative to the other observations in the sample. (Think about a dummy regressor which takes the value 1 for only one observation in the sample.)
- Properties:
  1. $0 \leq h_{ii} \leq 1$.
  2. $h_{ii} \geq 1/n$ if X includes an intercept.
  3. $\sum_{i=1}^{n} h_{ii} = k$.

Leave-One-Out Regression

- There are a number of statistical procedures – residual analysis, jackknife variance estimation, cross-validation, two-step estimation, hold-out sample evaluation – which make use of estimators constructed on sub-samples.
- Of particular importance is the case where we exclude a single observation and then repeat this for all observations. This is called **leave-one-out (LOO)** regression.

扔掉第i个观测值

Leave-One-Out Regression

- Specifically, the leave-one-out estimator of the regression coefficient β is the least squares estimator constructed using the full sample excluding a single observation i.

$$
\begin{aligned}
\widehat{\beta}_{(-i)} &= \left(\sum_{j \neq i} X_j X_j'\right)^{-1} \left(\sum_{j \neq i} X_j Y_j\right) \\
&= \left(\boldsymbol{X}'\boldsymbol{X} - X_i X_i'\right)^{-1} \left(\boldsymbol{X}'\boldsymbol{Y} - X_i Y_i\right) \\
&= \left(\boldsymbol{X}_{(-i)}'\boldsymbol{X}_{(-i)}\right)^{-1} \boldsymbol{X}_{(-i)}'\boldsymbol{Y}_{(-i)}
\end{aligned}
$$

- The leave-one-out predicted value for Y_i is $\widetilde{Y}_i = X_i'\widehat{\beta}_{(-i)}$.
- The leave-one-out residual, prediction error, or prediction residual is $\widetilde{e}_i = Y_i - \widetilde{Y}_i$.

Leave-One-Out Regression

- **Theorem:** The leave-one-out estimator and prediction error equal $\widehat{\beta}_{(-i)} = \widehat{\beta} - (\boldsymbol{X}'\boldsymbol{X})^{-1} X_i \widetilde{e}_i$ and $\widetilde{e}_i = (1 - h_{ii})^{-1} \widehat{e}_i$.
- Define

$$\begin{aligned}\boldsymbol{M}^* &= (\boldsymbol{I}_n - \operatorname{diag}\{h_{11}, .., h_{nn}\})^{-1} \\ &= \operatorname{diag}\left\{(1 - h_{11})^{-1}, .., (1 - h_{nn})^{-1}\right\}\end{aligned}$$

  Then $\widetilde{\boldsymbol{e}} = \boldsymbol{M}^* \widehat{\boldsymbol{e}}$
- One use of the prediction errors is to estimate the out-of-sample mean squared error:

$$\widetilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \widetilde{e}_i^2 = \frac{1}{n} \sum_{i=1}^{n} (1 - h_{ii})^{-2} \widehat{e}_i^2$$

- This is known as the sample mean squared prediction error.

Leave-One-Out Regression

- Prove the theorem:

$$\widehat{\beta}_{(-i)} = \left(\boldsymbol{X}'\boldsymbol{X} - X_i X_i'\right)^{-1} \left(\boldsymbol{X}'\boldsymbol{Y} - X_i Y_i\right)$$

- Multiply by $\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \left(\boldsymbol{X}'\boldsymbol{X} - X_i X_i'\right)$. We obtain

$$\begin{aligned}\widehat{\beta}_{(-i)} - \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} X_i X_i' \widehat{\beta}_{(-i)} &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \left(\boldsymbol{X}'\boldsymbol{Y} - X_i Y_i\right) \\ &= \widehat{\beta} - \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} X_i Y_i\end{aligned}$$

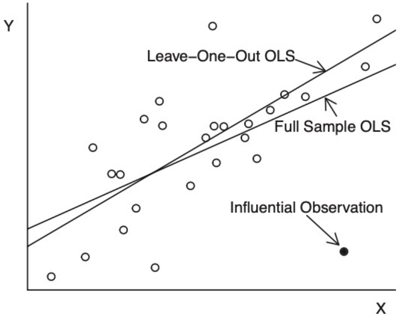
- Rewriting

$$\widehat{\beta}_{(-i)} = \widehat{\beta} - \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} X_i \left(Y_i - X_i' \widehat{\boldsymbol{\beta}}_{(-i)}\right) = \widehat{\beta} - \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} X_i \widetilde{e}_i$$

- Then $X_i' \widehat{\beta}_{(-i)} = X_i' \widehat{\beta} - X_i' \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} X_i \widetilde{e}_i = X_i' \widehat{\beta} - h_{ii} \widetilde{e}_i$
- Then we obtain $\widetilde{e}_i = \widehat{e}_i + h_{ii} \widetilde{e}_i$

Influential Observations

- Another use of the leave-one-out estimator is to investigate the impact of influential observations, sometimes called **outliers**.
- We say that observation i is influential if its omission from the sample induces a substantial change in a parameter estimate of interest. $\widehat{\beta} - \widehat{\beta}_{(-i)} = (\boldsymbol{X}'\boldsymbol{X})^{-1} X_i \widetilde{e}_i$



Influential Observations

- Two way to measure the impact of influential observation:

$$\widehat{\beta} - \widehat{\beta}_{(-i)} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} X_i \widetilde{e}_i$$

$$\widehat{Y}_i - \widetilde{Y}_i = X_i'\widehat{\beta} - X_i'\widehat{\beta}_{(-i)} = X_i' \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} X_i \widetilde{e}_i = h_{ii}\widetilde{e}_i$$

- Two indices:
  1. DFBETA: The ratio of the changes of coefficient to the coefficient's standard error
  2. Influence $= \max_{1 \leq i \leq n} \left| \widehat{Y}_i - \widetilde{Y}_i \right| = \max_{1 \leq i \leq n} \left| h_{ii}\widetilde{e}_i \right|$
- If an observation is determined to be influential what should be done?
  1. Take a close look at the data, to see whether the observation is incorrectly measured;
  2. If it is measured correctly, you can use other specification to model the observation or just delete it. However, the way you deal with data should be reasonable and transparent.