Advanced Econometrics

Lecture 3: The Algebra of Least Squares (Hansen Chapter 3)

Instructor: Ma, Jun

Renmin University of China

Fall 2018

样本

Samples

argmin$_{b \in \mathbb{R}^k}$ $E((Y - X'b)^2)$

$= (E(XX'))^{-1} E(XY)$

- Consider the best linear predictor of Y given $\boldsymbol{X}$ for a pair of random variables $(Y, \boldsymbol{X}) \in \mathbb{R} \times \mathbb{R}^k$ with joint distribution F and call this the linear projection model. We are interested in estimating the projection coefficients

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

- The dataset is $\{(Y_i, \boldsymbol{X}_i) : i = 1, \ldots, n\}$. We call this the **sample** or the **observations**.

$(Y_i, X_i) \sim F$

- From the viewpoint of empirical analysis, a dataset is an array of numbers often organized as a table, where the columns of the table correspond to distinct variables and the rows correspond to distinct observations.
- For empirical analysis, the dataset and observations are fixed in the sense that they are numbers presented to the researcher. For statistical analysis we need to view the dataset as random, or more precisely as a realization of a random process.

Samples

Assumption

The observations $\{(y_1, \boldsymbol{x}_1), \ldots, (y_i, \boldsymbol{x}_i), \ldots (y_n, \boldsymbol{x}_n)\}$ *are identically distributed; they are draws from a common distribution* F.

- In econometric theory, we refer to the underlying common distribution as the population. Some authors prefer the label the **data-generating-process** (DGP).
- In contrast we refer to the observations available to us $\{(Y_i, \boldsymbol{X}_i) : i = 1, ..., n\}$ as the sample or dataset.

Samples

We can write the model as

$$Y_i = \boldsymbol{X}_i' \beta + e_i,$$

where the linear projection coefficient $\boldsymbol{\beta}$ is defined as

$$\boldsymbol{\beta} = \underset{b \in \mathbb{R}^k}{\operatorname{argmin}} S(\boldsymbol{b}),$$

the minimizer of the expected squared error

$$S(\boldsymbol{b}) = \mathbb{E}\left(\left(Y_i - \boldsymbol{X}_i' \boldsymbol{b}\right)^2\right), \qquad = (\mathbb{E} X_i X_i')^{-1} \mathbb{E}(X_i Y_i)$$

and has the explicit solution

$$\boldsymbol{\beta} = \left(\mathbb{E}\left(\boldsymbol{X}_i \boldsymbol{X}_i'\right)\right)^{-1} \mathbb{E}\left(\boldsymbol{X}_i Y_i\right).$$

矩估计

Moment Estimators

- Suppose that we are interested in the population mean μ of a random variable Y_i with distribution function F

$$\mu = \mathbb{E}(Y_i) = \int_{-\infty}^{\infty} y dF(y) = \int_{-\infty}^{\infty} y f(y)\, dy.$$

总体期望

- The mean μ is a function of the distribution F. To estimate μ given a sample $\{Y_1, \ldots, Y_n\}$ a natural estimator is the sample mean:

$$\hat{\mu} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

$\hat{\mu}$ 无偏，μ 的矩估计

- Now suppose that we are interested in a set of population means of possibly non-linear functions of a random vector $\boldsymbol{Y}$, say $\boldsymbol{\mu} = \mathbb{E}(\boldsymbol{h}(\boldsymbol{Y}_i))$. For example, we may be interested in the first two moments of Y_i, $\mathbb{E}(Y_i)$ and $\mathbb{E}\left(Y_i^2\right)$. In this case the natural estimator is the vector of sample means,

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{h}(\boldsymbol{Y}_i).$$

For example, $\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} Y_i$ and $\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$. We call $\hat{\boldsymbol{\mu}}$ the moment estimator for $\boldsymbol{\mu}$.

Moment Estimators

- Now suppose that we are interested in a nonlinear function of a set of moments. For example,
$$\sigma^2 = \mathrm{Var}(Y_i) = \mathbb{E}\left(Y_i^2\right) - \left(\mathbb{E}(Y_i)\right)^2 .$$
Many parameters of interest can be written as a function of moments of $\boldsymbol{Y}$:
$$\boldsymbol{\beta} = \boldsymbol{g}(\boldsymbol{\mu}), \text{ where } \boldsymbol{\mu} = \mathbb{E}(\boldsymbol{h}(\boldsymbol{Y}_i)).$$
- In this context a natural estimator of $\boldsymbol{\beta}$ is obtained by replacing $\boldsymbol{\mu}$ with $\hat{\boldsymbol{\mu}}$:
$$\hat{\boldsymbol{\beta}} = \boldsymbol{g}(\hat{\boldsymbol{\mu}}), \text{ where } \hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{h}(\boldsymbol{Y}_i) .$$
We call $\hat{\boldsymbol{\beta}}$ a moment estimator of $\boldsymbol{\beta}$. For example, the moment estimator of σ^2 is
$$\hat{\sigma}^2 = \hat{\mu}_2 - \hat{\mu}_1^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2 .$$

最小二乘估计

Least Squares Estimator

- The moment estimator of $S(\boldsymbol{b})$ is the sample average:

$$\widehat{S}(\boldsymbol{b}) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \boldsymbol{X}_i'\boldsymbol{b}\right)^2$$
$$= \frac{1}{n} SSE(\boldsymbol{b})$$

对应每一个b的残差平方和.

where

$$SSE(\boldsymbol{b}) = \sum_{i=1}^{n}\left(Y_i - \boldsymbol{X}_i'\boldsymbol{b}\right)^2$$

is called the sum-of-squared-errors function.

- Since the projection coefficient minimizes $S(\boldsymbol{b})$, the OLS estimator minimizes $\widehat{S}(\boldsymbol{b})$:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{b}\in\mathbb{R}^k}{\operatorname{argmin}}\, \hat{S}(\boldsymbol{b}) = \underset{\boldsymbol{b}\in\mathbb{R}^k}{\operatorname{argmin}}\, SSE(\boldsymbol{b}).$$

Solving for Least Squares with One Regressor

- Consider the case $k = 1$ so that the coefficient β is a scalar. Then

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

- The minimizer of $SSE(\beta)$ is

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

- The intercept-only model: $X_i = 1$ and

$$\hat{\beta} = \frac{\sum_{i=1}^{n} 1 Y_i}{\sum_{i=1}^{n} 1^2} = \frac{1}{n} \sum_{i=1}^{n} Y_i = \bar{Y}.$$

Solving for Least Squares with Multiple Regressors

- Expand SSE to find

$$SSE(\boldsymbol{b}) = \sum_{i=1}^{n} Y_i^2 - 2\boldsymbol{b}' \sum_{i=1}^{n} \boldsymbol{X}_i Y_i + \boldsymbol{b}' \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i' \boldsymbol{b}.$$

- The first-order condition is

用矩阵微积分解一阶条件. $\Rightarrow$

$$0 = \frac{\partial}{\partial \boldsymbol{b}} SSE\left(\hat{\boldsymbol{\beta}}\right) = -2 \sum_{i=1}^{n} \boldsymbol{X}_i Y_i + 2 \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i' \hat{\boldsymbol{\beta}},$$

which is actually a system of k equations with k unknowns.

- We find an explicit formula for the OLS:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i' \right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{X}_i Y_i \right).$$

最小二乘的矩阵表述

Solving for Least Squares with Multiple Regressors

- Alternatively, we can write the projection coefficient $\boldsymbol{\beta}$ as an explicit function of the moments $\boldsymbol{Q}_{\boldsymbol{XY}} = \mathbb{E}(\boldsymbol{X}Y)$ and $\boldsymbol{Q}_{\boldsymbol{XX}} = \mathbb{E}(\boldsymbol{X}\boldsymbol{X}')$. Their moment estimators are

$$\hat{\boldsymbol{Q}}_{\boldsymbol{XY}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i Y_i \text{ and } \hat{\boldsymbol{Q}}_{\boldsymbol{XX}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i'.$$

- The moment estimator of $\boldsymbol{\beta}$ replaces the population moments with the sample moments:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \hat{\boldsymbol{Q}}_{\boldsymbol{XX}}^{-1} \hat{\boldsymbol{Q}}_{\boldsymbol{XY}} \\ &= \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i Y_i\right) \\ &= \left(\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i'\right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{X}_i Y_i\right) \end{aligned}$$

矩估计和最小二乘估计是一样的.

which is identical with OLS.

Solving for Least Squares with Multiple Regressors

Definitions

The **Least-squares estimator $\hat{\boldsymbol{\beta}}$ is**

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^{\mathrm{k}}}{\operatorname{argmin}} \hat{\mathrm{S}}\left(\boldsymbol{\beta}\right)$$

where

$$\hat{S}\left(\boldsymbol{\beta}\right) = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \boldsymbol{X}_i' \boldsymbol{\beta}\right)^2$$

and has the solution

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i'\right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{X}_i Y_i\right).$$

最小二乘残差

Least Squares Residuals

- Define the fitted value $\hat{Y}_i = \boldsymbol{X}_i'\hat{\boldsymbol{\beta}}$ and the residual
$$\hat{e}_i = Y_i - \hat{Y}_i = Y_i - \boldsymbol{X}_i'\hat{\boldsymbol{\beta}}.$$
- Note that $Y_i = \hat{Y}_i + \hat{e}_i$ and $Y_i = \boldsymbol{X}_i'\hat{\boldsymbol{\beta}} + \hat{e}_i$.
- e_i is called error and $\hat{e}_i$ is called residual. The OLS first-order condition implies $\sum_{i=1}^n \boldsymbol{X}_i\hat{e}_i = \boldsymbol{0}$.
- Alternatively,

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

- When $\boldsymbol{X}_i$ contains a constant, $\frac{1}{n}\sum_{i=1}^n \hat{e}_i = 0$. Thus the residuals have a sample mean of zero and the sample correlation between the regressors and the residual is zero.

定义拟合值 $\hat{Y}_i$

残差 $Y_i - \hat{Y}_i = \hat{e}_i$

在样本中 $\hat{e}$和 X 是正交的.

$\sum_{i=1}^n \underbrace{X_i}_{k\times 1} \hat{e}_i = 0$

$(X_{1i}, X_{2i}, \cdots, X_{ki}) \begin{pmatrix} \hat{e}_1 \\ \hat{e}_2 \\ \vdots \\ \hat{e}_n \end{pmatrix} = \vec{0}$

$\sum_{i=1}^n X_{ji}\hat{e}_i = 0, \ \forall j = 1, \cdots, k.$

如果 X_i 中包含一个常数，拟合残差的平均值为 0.

被解释变量

Demeaned Regressors

- Sometimes it is useful to separate the constant from the other regressors, and write the linear projection equation in the format

$$Y_i = \boldsymbol{X}_i'\boldsymbol{\beta} + \alpha + e_i$$

where α is the intercept and $\boldsymbol{X}_i$ does not contain a constant.

- The least-squares estimates and residuals can be written as

$$Y_i = \boldsymbol{X}_i'\hat{\boldsymbol{\beta}} + \hat{\alpha} + \hat{e}_i.$$

Then $\sum_{i=1}^n \boldsymbol{X}_i\hat{e}_i = \boldsymbol{0}$ can be written as

$$\sum_{i=1}^n \left(Y_i - \boldsymbol{X}_i'\hat{\boldsymbol{\beta}} - \hat{\alpha}\right) = 0 \text{ and } \sum_{i=1}^n \boldsymbol{X}_i\left(Y_i - \boldsymbol{X}_i'\hat{\boldsymbol{\beta}} - \hat{\alpha}\right) = \boldsymbol{0}.$$

- Inserting $\hat{\alpha} = \bar{Y} - \bar{\boldsymbol{X}}'\hat{\boldsymbol{\beta}}$ into the second equation:

$$\sum_{i=1}^n \boldsymbol{X}_i\left(\left(Y_i - \bar{Y}\right) - \left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}\right)'\hat{\boldsymbol{\beta}}\right) = 0.$$

- Solving for $\hat{\boldsymbol{\beta}}$ we find

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^n \left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}\right)'\right)^{-1}\left(\sum_{i=1}^n \left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}\right)\left(Y_i - \bar{Y}\right)\right).$$

XX' $\quad XY_{k\times 1}$

Model in Matrix Notation

- We can stack these n equations together

$$
\begin{aligned}
Y_1 &= \boldsymbol{X}_1'\boldsymbol{\beta} + e_1 \\
Y_2 &= \boldsymbol{X}_2'\boldsymbol{\beta} + e_2 \\
&\vdots \\
Y_n &= \boldsymbol{X}_n'\boldsymbol{\beta} + e_n.
\end{aligned}
$$

- Define

$$
\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1' \\ \boldsymbol{X}_2' \\ \vdots \\ \boldsymbol{X}_n' \end{pmatrix}, \quad \boldsymbol{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}.
$$

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

- The system of n equations can be written as

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

$$
\underset{k\times 1}{\hat{\beta}} = \Big(\sum_{i=1}^{n} \underset{k\times k}{X_i X_i'}\Big)\Big(\sum_{i=1}^{n} X_i Y_i\Big)
$$

$$
X_i = \begin{pmatrix} X_{i,1} \\ \vdots \\ X_{i,k} \end{pmatrix}_{k\times 1}
$$

i的观测值

$$
Y_i = \beta_1 X_{1,i} + \cdots + \beta_k X_{k,i} + e_i, \quad i = 1, \dots, n
$$

$$
= X_i'\beta + e_i \leftarrow \text{the model}
$$

$$
\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}
$$

$$
X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} X_{1,1} & \cdots & X_{1,k} \\ \vdots & & \vdots \\ X_{n,1} & \cdots & X_{n,k} \end{pmatrix}_{n\times k}, \quad Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}_{n\times 1}, \quad e = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}_{n\times 1}
$$

$$
\begin{aligned}
Y_1 &= X_1'\beta + e_1 \\
Y_2 &= X_2'\beta + e_2 \\
&\vdots \\
Y_n &= X_n'\beta + e_n
\end{aligned}
\longrightarrow \underset{n\times 1}{Y} = \underset{n\times k}{X}\,\underset{k\times 1}{\beta} + \underset{n\times 1}{e}
$$

Model in Matrix Notation

一个重要假设: $X'X$ 是可逆的, $n>k$.

$$\hat{\beta}=\Big(\underbrace{\sum_{i=1}^{n} X_i X_i'}_{X'X}\Big)^{-1}\Big(\underbrace{\sum_{i=1}^{n} X_i Y_i}_{X'Y}\Big)$$

$$\hat{\beta}=(\underset{k\times n}{X'}\ \underset{n\times k}{X})^{-1}(X'Y)$$

- Sample sums can be written in matrix notation:

$$\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i' = \boldsymbol{X}'\boldsymbol{X} \text{ and } \sum_{i=1}^{n} \boldsymbol{X}_i Y_i = \boldsymbol{X}'\boldsymbol{Y}.$$

- Therefore the least-squares estimator can be written as

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{Y}\right).$$ ← X矩阵满秩, 无多重共线性假设, $n>k$.

- The residual vector is $\hat{\boldsymbol{e}} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$. We can write $\sum_{i=1}^{n} \boldsymbol{X}_i \hat{e}_i = \boldsymbol{0}$ as $\boldsymbol{X}'\hat{\boldsymbol{e}} = \boldsymbol{0}$.

回归残差 $\hat{e}$

$$\sum_{i=1}^{n} \underset{k\times 1}{X_i}\,\hat{e}_i = 0 \rightarrow X'\hat{e}=0$$

Model in Matrix Notation

Important Matrix Expressions

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

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

$$\hat{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$$

$$\boldsymbol{X}'\hat{\boldsymbol{e}} = \boldsymbol{0}$$

投影矩阵

Projection Matrix

拟合值 $\hat{Y} = X\hat{\beta} = \underbrace{X(X'X)^{-1}X'}_{P_{n\times n} \text{投影矩阵}} Y$

- Define
$$\boldsymbol{P} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'.$$
Observe
$$\boldsymbol{P}\boldsymbol{X} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{X}.$$
This is a property of a **projection** matrix.

如果 $X: n\times k$. $k<n$

那么 $P\vec{y}$ 就是 $\vec{y}$ 在X的列空间上的投影.

- For any matrix $\boldsymbol{Z}$ which can be written as $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{\Gamma}$ for some matrix $\boldsymbol{\Gamma}$,
$$\boldsymbol{P}\boldsymbol{Z} = \boldsymbol{P}\boldsymbol{X}\boldsymbol{\Gamma} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\Gamma} = \boldsymbol{X}\boldsymbol{\Gamma} = \boldsymbol{Z}.$$
- If we partition the matrix $\boldsymbol{X}$ into two matrices $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ so that
$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 \end{bmatrix}$$
then $\boldsymbol{P}\boldsymbol{X}_1 = \boldsymbol{X}_1$.

把矩阵X分块.

$PX = P[X_1 \ \ X_2] = [PX_1 \ \ PX_2]$

Projection Matrix

The matrix $\boldsymbol{P}$ is symmetric and idempotent:

$$\begin{aligned} \boldsymbol{P}' &= \left(\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right)' \\ &= \left(\boldsymbol{X}'\right)'\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right)'\left(\boldsymbol{X}\right)' \\ &= \boldsymbol{X}\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)'\right)^{-1}\boldsymbol{X}' \\ &= \boldsymbol{X}\left(\left(\boldsymbol{X}\right)'\left(\boldsymbol{X}'\right)'\right)^{-1}\boldsymbol{X}' \\ &= \boldsymbol{P} \end{aligned}$$

运用 $(A^{-1})' = (A')^{-1}$

$\Rightarrow P' = P$

P是对称的

$$\begin{aligned} \boldsymbol{P}\boldsymbol{P} &= \boldsymbol{P}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' \\ &= \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' \\ &= \boldsymbol{P}. \end{aligned}$$

P是幂等的

Projection Matrix

- The matrix $\boldsymbol{P}$ has the property that it creates the fitted values in a least-squares regression:

$$\boldsymbol{P}\boldsymbol{y} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{Y}}.$$

- A special example of a projection matrix occurs when $\boldsymbol{X} = \mathbf{1}$ is an n-vector of ones. Then

$$\begin{aligned}\boldsymbol{P}_1 &= \mathbf{1}\left(\mathbf{1}'\mathbf{1}\right)^{-1}\mathbf{1}' \\ &= \frac{1}{n}\mathbf{1}\mathbf{1}'.\end{aligned}$$

$$P_1 Y = \mathbf{1}(\underbrace{\mathbf{1}'\mathbf{1}}_{=n})^{-1}\underbrace{\mathbf{1}'Y}_{\sum_{i=1}^{n} Y_i}$$

$$= \frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y} = \begin{pmatrix}\bar{Y}\\ \vdots \\ \bar{Y}\end{pmatrix}_{n\times 1}$$

- Note

$$\begin{aligned}\boldsymbol{P}_1\boldsymbol{y} &= \mathbf{1}\left(\mathbf{1}'\mathbf{1}\right)^{-1}\mathbf{1}'\boldsymbol{Y} \\ &= \mathbf{1}\bar{Y}\end{aligned}$$

creates an n-vector whose elements are the sample mean $\bar{Y}$ of Y_i

Projection Matrix

投影矩阵是对称的,幂等的,但不是满秩的.

$\text{rank}(P) = k.$

Theorem

The i^{th} diagonal element of $\boldsymbol{P} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'$ *is*

$$h_{ii} = \boldsymbol{X}_i'\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}_i.$$

$$\sum_{i=1}^{n} h_{ii} = \text{tr}\boldsymbol{P} = k$$

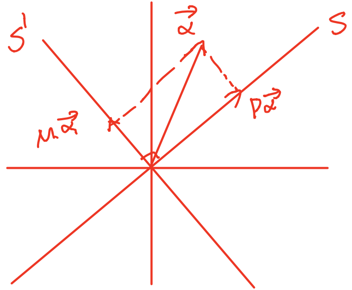
and $0 \leq h_{ii} \leq 1$.

$$\begin{aligned}
\text{tr}\boldsymbol{P} &= \text{tr}\left(\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right) \\
&= \text{tr}\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\right) \\
&= \text{tr}\left(\boldsymbol{I}_k\right) \\
&= k
\end{aligned}$$

One implication is that the rank of $\boldsymbol{P}$ is k.

正交投影矩阵

Orthogonal Projection



S与S'是正交空间.

- Define

$$M = I_n - P$$
$$= I_n - X\left(X'X\right)^{-1} X'.$$

- Note

矩阵M与X的所有列都是正交的.

$$MX = (I_n - P)\,X = X - PX = X - X = 0.$$

Thus M and X are orthogonal. We call M the **orthogonal projection matrix**.

- If $Z = X\Gamma$, then

$$MZ = Z - PZ = 0.$$

For example, $MX_1 = 0$ for any subcomponent X_1 of X and $MP = 0$.

- M is symmetric ($M' = M$) and idempotent ($MM = M$). tr$M = n - k$. The rank of M is $n - k$.

对称. 幂等.

rank(M) = n-k   trM = n-k

Orthogonal Projection

- $\boldsymbol{M}$ creates least-square residuals:

$$\boldsymbol{MY} = \boldsymbol{Y} - \boldsymbol{PY} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{e}}.$$

$\hat{e} = Y - X\hat{\beta} = Y - \hat{Y} = Y - PY$
$= (I_n - P)Y = MY$

- When $\boldsymbol{X} = \mathbf{1}$,

$$\begin{aligned} \boldsymbol{M}_1 &= \boldsymbol{I}_n - \boldsymbol{P}_1 \\ &= \boldsymbol{I}_n - \mathbf{1}\left(\mathbf{1}'\mathbf{1}\right)^{-1}\mathbf{1}' \end{aligned}$$

$M_1 = Y - \begin{pmatrix} \bar{Y} \\ \vdots \\ \bar{Y} \end{pmatrix} = \begin{pmatrix} Y_1 - \bar{Y} \\ \vdots \\ Y_n - \bar{Y} \end{pmatrix}$

and $\boldsymbol{M}_1$ creates demeaned values $\boldsymbol{M}_1\boldsymbol{Y} = \boldsymbol{Y} - \mathbf{1}\bar{Y}$.

$Y_i - \bar{Y}$ — 离差.

- We find

$$\hat{\boldsymbol{e}} = \boldsymbol{MY} = \boldsymbol{M}(\boldsymbol{X\beta} + \boldsymbol{e}) = \boldsymbol{Me}.$$

误差估计

Estimation of Error Variance

- If e_i were observed, we would estimate $\sigma^2 = \mathbb{E}\left(e_i^2\right)$ by
$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 .$$
This is infeasible as e_i is not observed.
- The feasible estimator: $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{e}_i^2$. In matrix notation,
$$\tilde{\sigma}^2 = n^{-1}\boldsymbol{e}'\boldsymbol{e} \text{ and } \hat{\sigma}^2 = n^{-1}\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}}.$$
- Since $\hat{\boldsymbol{e}} = \boldsymbol{MY} = \boldsymbol{Me}$,
$$\begin{aligned}\hat{\sigma}^2 &= n^{-1}\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}} \\ &= n^{-1}\boldsymbol{Y}'\boldsymbol{MMY} \\ &= n^{-1}\boldsymbol{Y}'\boldsymbol{MY} \\ &= n^{-1}\boldsymbol{e}'\boldsymbol{Me}.\end{aligned}$$
- An implication:
$$\begin{aligned}\tilde{\sigma}^2 - \hat{\sigma}^2 &= n^{-1}\boldsymbol{e}'\boldsymbol{e} - n^{-1}\boldsymbol{e}'\boldsymbol{Me} \\ &= n^{-1}\boldsymbol{e}'\boldsymbol{Pe} \\ &\geq 0.\end{aligned}$$

$\tilde{\sigma}^2$ — 不可行估计（误差项均值）

$\hat{\sigma}^2$ — 可行估计（残差均值）

Analysis of Variance

- Write
$$\boldsymbol{Y} = \boldsymbol{PY} + \boldsymbol{MY} = \hat{\boldsymbol{Y}} + \hat{\boldsymbol{e}},$$
where
$$\hat{\boldsymbol{Y}}'\hat{\boldsymbol{e}} = (\boldsymbol{PY})'(\boldsymbol{MY}) = \boldsymbol{Y}'\boldsymbol{PMY} = 0.$$
- Then
$$\boldsymbol{Y}'\boldsymbol{Y} = \hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}} + 2\hat{\boldsymbol{Y}}'\hat{\boldsymbol{e}} + \hat{\boldsymbol{e}}'\hat{\boldsymbol{e}} = \hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}} + \hat{\boldsymbol{e}}'\hat{\boldsymbol{e}}$$
or
$$\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} \hat{Y}_i^2 + \sum_{i=1}^{n} \hat{e}_i^2.$$

Analysis of Variance

- Since $\boldsymbol{Y} = \hat{\boldsymbol{Y}} + \hat{\boldsymbol{e}}$,

$$\boldsymbol{Y} - \underbrace{\mathbf{1}\bar{Y}}_{=\begin{pmatrix}\bar{Y}\\ \vdots \\ \bar{Y}\end{pmatrix}_{n\times 1}} = \hat{\boldsymbol{Y}} - \mathbf{1}\bar{Y} + \hat{\boldsymbol{e}},$$

  where

$$\left(\hat{\boldsymbol{Y}} - \mathbf{1}\bar{Y}\right)' \hat{\boldsymbol{e}} = \underbrace{\hat{\boldsymbol{Y}}'\hat{\boldsymbol{e}}}_{=0} - \bar{Y}\underbrace{\mathbf{1}'\hat{\boldsymbol{e}}}_{=0} = 0.$$

  $= 0 \rightarrow$ if 回归中有截距项

- Then

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

  or

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

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

$$\mathbf{1}'Y = \mathbf{1}'\hat{Y} + \underbrace{\mathbf{1}'\hat{e}}_{=0 \text{ if 有截距项}}$$

$$\frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n}\hat{Y}_i$$

当回归中有截距项时，样本均值等于样本估计量的均值.

Analysis of Variance

- A commonly reported statistic is the R-squared:

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

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

回归平方和 (SSR) + 残差平方和 (SSE)

= 总离差平方和 (SST)

  This is a measure of goodness of regression 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.

只要加入新的变量，R^2 一定变大

回归成分

Regression Components

- Partition

$$\underset{n\times k}{\boldsymbol{X}} = \left[\ \underset{n\times k_1}{\boldsymbol{X}_1} \quad \underset{n\times k_2}{\boldsymbol{X}_2}\ \right]$$

and

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} \begin{matrix} k_1\times 1 \\ \cdot \\ k_2\times 2 \end{matrix}$$

- Then the regression model can be rewritten as

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

and

$$\boldsymbol{Y} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{e}} = \boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_2\hat{\boldsymbol{\beta}}_2 + \hat{\boldsymbol{e}}.$$

$$Y = X\beta + e = X_1\beta_1 + X_2\beta_2 + e$$

$$\hat{Y} = X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \hat{e}$$

$$X'X = \begin{bmatrix} X_1' \\ X_2' \end{bmatrix}\begin{bmatrix} X_1 & X_2 \end{bmatrix} = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} X_1'Y \\ X_2'Y \end{bmatrix}$$

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X'X)^{-1}(X'Y) = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1}\begin{bmatrix} X_1'Y \\ X_2'Y \end{bmatrix}$$

Regression Components

- Partition

$$\hat{Q}_{XX} = \begin{bmatrix} \hat{Q}_{11} & \hat{Q}_{12} \\ \hat{Q}_{21} & \hat{Q}_{22} \end{bmatrix} = \begin{bmatrix} \frac{1}{n}X_1'X_1 & \frac{1}{n}X_1'X_2 \\ \frac{1}{n}X_2'X_1 & \frac{1}{n}X_2'X_2 \end{bmatrix}; \hat{Q}_{XY} = \begin{bmatrix} \hat{Q}_{1Y} \\ \hat{Q}_{2Y} \end{bmatrix} = \begin{bmatrix} \frac{1}{n}X_1'Y \\ \frac{1}{n}X_2'Y \end{bmatrix}.$$

- By the partitioned matrix inversion,

$$\hat{Q}_{XX}^{-1} = \begin{bmatrix} \hat{Q}_{11} & \hat{Q}_{12} \\ \hat{Q}_{21} & \hat{Q}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \hat{Q}^{11} & \hat{Q}^{12} \\ \hat{Q}^{21} & \hat{Q}^{22} \end{bmatrix} = \begin{bmatrix} \hat{Q}_{11\cdot 2}^{-1} & -\hat{Q}_{11\cdot 2}^{-1}\hat{Q}_{12}\hat{Q}_{22}^{-1} \\ -\hat{Q}_{22\cdot 1}^{-1}\hat{Q}_{21}\hat{Q}_{11}^{-1} & \hat{Q}_{22\cdot 1}^{-1} \end{bmatrix},$$

where $\hat{Q}_{11\cdot 2} = \hat{Q}_{11} - \hat{Q}_{12}\hat{Q}_{22}^{-1}\hat{Q}_{21}$ and $\hat{Q}_{22\cdot 1} = \hat{Q}_{22} - \hat{Q}_{21}\hat{Q}_{11}^{-1}\hat{Q}_{12}$.

- Thus,

$$\begin{aligned} \hat{\beta} &= \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \\ &= \begin{bmatrix} \hat{Q}_{11\cdot 2}^{-1} & -\hat{Q}_{11\cdot 2}^{-1}\hat{Q}_{12}\hat{Q}_{22}^{-1} \\ -\hat{Q}_{22\cdot 1}^{-1}\hat{Q}_{21}\hat{Q}_{11}^{-1} & \hat{Q}_{22\cdot 1}^{-1} \end{bmatrix} \begin{bmatrix} \hat{Q}_{1Y} \\ \hat{Q}_{2Y} \end{bmatrix} \\ &= \begin{pmatrix} \hat{Q}_{11\cdot 2}^{-1}\hat{Q}_{1Y\cdot 2} \\ \hat{Q}_{22\cdot 1}^{-1}\hat{Q}_{2Y\cdot 1} \end{pmatrix}, \end{aligned}$$

where $\hat{Q}_{1Y\cdot 2} = \hat{Q}_{1Y} - \hat{Q}_{12}\hat{Q}_{22}^{-1}\hat{Q}_{2Y}$.

$$\begin{bmatrix} \hat{Q}^{11} & \hat{Q}^{12} \\ \hat{Q}^{21} & \hat{Q}^{22} \end{bmatrix}$$

$$\Rightarrow \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \hat{Q}^{11}X_1'Y + \hat{Q}^{12}X_2'Y \\ \hat{Q}^{21}X_1'Y + \hat{Q}^{22}X_2'Y \end{pmatrix} = \begin{pmatrix} (\hat{Q}^{11}X_1' + \hat{Q}^{12}X_2')Y \\ (\hat{Q}^{21}X_1' + \hat{Q}^{22}X_2')Y \end{pmatrix}$$

$$\underbrace{\hat{Q}^{11}X_1' + \hat{Q}^{12}X_2'}_{(X_1'M_2X_1)^{-1}X_1'M_2}$$

Regression Components

- Now
$$\begin{aligned}\hat{\boldsymbol{Q}}_{11\cdot 2} &= \hat{\boldsymbol{Q}}_{11} - \hat{\boldsymbol{Q}}_{12}\hat{\boldsymbol{Q}}_{22}^{-1}\hat{\boldsymbol{Q}}_{21} \\ &= \frac{1}{n}\boldsymbol{X}_1'\boldsymbol{X}_1 - \frac{1}{n}\boldsymbol{X}_1'\boldsymbol{X}_2\left(\frac{1}{n}\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\frac{1}{n}\boldsymbol{X}_2'\boldsymbol{X}_1 \\ &= \frac{1}{n}\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\end{aligned}$$
where
$$\boldsymbol{M}_2 = \boldsymbol{I}_n - \boldsymbol{X}_2\left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'$$
is the orthogonal projection matrix for $\boldsymbol{X}_2$.
- Also
$$\begin{aligned}\hat{\boldsymbol{Q}}_{1Y\cdot 2} &= \hat{\boldsymbol{Q}}_{1Y} - \hat{\boldsymbol{Q}}_{12}\hat{\boldsymbol{Q}}_{22}^{-1}\hat{\boldsymbol{Q}}_{2Y} \\ &= \frac{1}{n}\boldsymbol{X}_1'\boldsymbol{Y} - \frac{1}{n}\boldsymbol{X}_1'\boldsymbol{X}_2\left(\frac{1}{n}\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\frac{1}{n}\boldsymbol{X}_2'\boldsymbol{Y} \\ &= \frac{1}{n}\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{Y}.\end{aligned}$$
- Therefore
$\hat{\boldsymbol{\beta}}_1 = \left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)^{-1}\left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{Y}\right)$ and similarly $\hat{\boldsymbol{\beta}}_2 = \left(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2\right)^{-1}\left(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{Y}\right)$.

残差回归

Residual Regression

- Note

$$\begin{aligned}\hat{\boldsymbol{\beta}}_2 &= \left(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2\right)^{-1}\left(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{Y}\right)\\ &= \left(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{M}_1\boldsymbol{X}_2\right)^{-1}\left(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{M}_1\boldsymbol{Y}\right)\\ &= \left(\underset{n\times k_2}{\widetilde{\boldsymbol{X}}_2'\widetilde{\boldsymbol{X}}_2}\right)^{-1}\left(\widetilde{\boldsymbol{X}}_2'\underset{n\times 1}{\tilde{\boldsymbol{e}}_1}\right)\end{aligned}$$

where

$$\widetilde{\boldsymbol{X}}_2 = \boldsymbol{M}_1\boldsymbol{X}_2 \text{ and } \tilde{\boldsymbol{e}}_1 = \boldsymbol{M}_1\boldsymbol{Y}.$$

- The estimate $\hat{\boldsymbol{\beta}}_2$ is algebraically equal to the least-squares regression of $\tilde{\boldsymbol{e}}_1$ on $\widetilde{\boldsymbol{X}}_2$. $\tilde{\boldsymbol{e}}_1$ is the least-squares residuals from a regression of $\boldsymbol{Y}$ on $\boldsymbol{X}_1$. The columns of $\widetilde{\boldsymbol{X}}_2$ are the least-squares residuals from the regressions of the columns of $\boldsymbol{X}_2$ on $\boldsymbol{X}_1$.

$(X_1'M_2X_1)^{-1}X_1'M_2Y$

$= (X_1'M_2X_1)^{-1}X_1'M_2(X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \hat{e})$

$= \underbrace{(X_1'M_2X_1)^{-1}X_1'M_2X_1}_{单位阵\ I_n}\hat{\beta}_1$

$+ (X_1'M_2X_1)^{-1}X_1'\underbrace{M_2X_2}_{=0}\hat{\beta}_2$

$+ (X_1'M_2X_1)^{-1}X_1'\underbrace{M_2\hat{e}}_{= M_2\overset{M}{\widetilde{M}}e = Me = \hat{e}}$

$X_1'\hat{e} = 0$

$\Rightarrow (X_1'M_2X_1)^{-1}X_1'M_2Y = \hat{\beta}_1$

Residual Regression

*最重要的内容

Theorem

Frisch-Waugh-Lovell (FWL)

The OLS estimator of $\boldsymbol{\beta}_2$ and the OLS residuals $\hat{\boldsymbol{e}}$ may be equivalently computed by either the OLS regression or via the following algorithm:

1. Regress $\boldsymbol{Y}$ on $\boldsymbol{X}_1$, obtain residuals $\tilde{\boldsymbol{e}}_1$;
2. Regress $\boldsymbol{X}_2$ on $\boldsymbol{X}_1$, obtain residuals $\widetilde{\boldsymbol{X}}_2$;
3. Regress $\tilde{\boldsymbol{e}}_1$ on $\widetilde{\boldsymbol{X}}_2$, obtain OLS estimates $\hat{\boldsymbol{\beta}}_2$ and residuals $\hat{\boldsymbol{e}}$.

- To check (3), note

$$
\begin{aligned}
&\left(\boldsymbol{I}_n - \widetilde{\boldsymbol{X}}_2\left(\widetilde{\boldsymbol{X}}_2'\widetilde{\boldsymbol{X}}_2\right)^{-1}\widetilde{\boldsymbol{X}}_2'\right)\tilde{\boldsymbol{e}}_1\\
=&\boldsymbol{M}_1\boldsymbol{Y} - \boldsymbol{M}_1\boldsymbol{X}_2\hat{\boldsymbol{\beta}}_2\\
=&\boldsymbol{M}_1\left(\boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_2\hat{\boldsymbol{\beta}}_2 + \hat{\boldsymbol{e}}\right) - \boldsymbol{M}_1\boldsymbol{X}_2\hat{\boldsymbol{\beta}}_2\\
=&\boldsymbol{M}_1\hat{\boldsymbol{e}}\\
=&\hat{\boldsymbol{e}}.
\end{aligned}
$$

FWL定理

$\hat\beta = (X'X)^{-1}(X'Y)$

$\Rightarrow \hat\beta_2 = (X_2'M_2X_2)^{-1}(X_2'M_2Y)$

最初作用是计算时降维.

回归 $Y \to X$: $\tilde e_1 = M_1 Y$

$X_2 \to X_1$: $\tilde X_2 = M_1 X_2$

$\tilde e_1 \to \tilde X_2$: $\hat\beta_2 = [(\tilde X_2'\tilde X_2)^{-1}\tilde X_2']\tilde e_1$

$\hat e = M_1Y - M_1X_2\hat\beta_2$

$= M_1(X_1\hat\beta_1 + X_2\hat\beta_2) + \hat e - M_1X_2\hat\beta_2$

$= I_n - \tilde X_2\hat\beta_2$

最小二乘法

$$\min_{b_0,b_1,\cdots,b_k} \sum_{i=1}^{n}\left(Y_i - b_0 - b_1 X_{1,i} - b_2 X_{2,i} - \cdots - b_k X_{k,i}\right)^2$$

$$Y_i - (1, X_{1,i} \cdots X_{k,i})' \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{pmatrix}$$

$$\begin{pmatrix} 1 & x_{1,1} & \cdots & x_{k,1} \\ 1 & x_{1,2} & \cdots & x_{k,2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1,n} & \cdots & x_{k,n} \end{pmatrix}$$

$$\downarrow$$

$$\min_{b} (Y - Xb)'(Y - Xb)$$

$$X' \underset{n\times 1}{\hat{e}} = 0$$

$$(1, 1, \cdots, 1)\hat{e} = \sum_{i=1}^{n} \hat{e}_i = 0$$

如果没有截距项， $X'\hat{e} = 0$ 成立

$$(1, 1, \cdots, 1)\hat{e} = \sum_{i=1}^{n} \hat{e}_i \neq 0$$ 回归残差和不一定等于0。