Advanced Econometrics

Lecture 5: The Algebra of Least Squares (Hansen Chapter 3)

Instructor: Ma, Jun

Renmin University of China

October 10, 2021

Samples

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

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.

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

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

and has the explicit solution

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

Moment estimators

- Suppose that we are interested in the population mean μ of a random variable Y_i: $\mu = \mathbb{E}(Y_i)$.
- The mean μ is a function of the distribution F. To estimate μ given a sample $\{Y_1, ..., Y_n\}$ a natural estimator is the sample mean $\hat{\mu} = \bar{Y} = n^{-1} \sum_{i=1}^{n} Y_i$.
- 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 $\mathbb{E}(Y_i^2)$. In this case the natural estimator is the vector of sample means, $\hat{\boldsymbol{\mu}} = n^{-1} \sum_{i=1}^{n} \boldsymbol{h}(\boldsymbol{Y}_i)$.
- For example, $\hat{\mu}_1 = n^{-1} \sum_{i=1}^{n} Y_i$ and $\hat{\mu}_2 = n^{-1} \sum_{i=1}^{n} Y_i^2$. We call $\hat{\boldsymbol{\mu}}$ the moment estimator for $\boldsymbol{\mu}$.

- Now suppose that we are interested in a nonlinear function of a set of moments. For example,

$$\sigma^2 = \text{Var}\left(Y_i\right) = \mathbb{E}\left(Y_i^2\right) - \left(\mathbb{E}\left(Y_i\right)\right)^2 .$$

Many parameters of interest can be written as a function of moments of $\boldsymbol{Y}$:

$$\boldsymbol{\beta} = \boldsymbol{g}\left(\boldsymbol{\mu}\right), \text{ where } \boldsymbol{\mu} = \mathbb{E}\left(\boldsymbol{h}\left(\boldsymbol{Y}_i\right)\right).$$

- In this context a natural estimator of $\boldsymbol{\beta}$ is obtained by replacing $\boldsymbol{\mu}$ with $\hat{\boldsymbol{\mu}}$:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{g}\left(\hat{\boldsymbol{\mu}}\right), \text{ where } \hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{h}\left(\boldsymbol{Y}_i\right).$$

We call $\hat{\boldsymbol{\beta}}$ a moment estimator of $\boldsymbol{\beta}$. For example, the moment estimator of σ^2 is

$$\hat{\sigma}^2 = \hat{\mu}_2 - \hat{\mu}_1^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2 .$$

Least squares estimator

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

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

  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

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

- 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

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

- The first-order condition is

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

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

- We find an explicit formula for the OLS:

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

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

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

  which is identical with OLS.

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, $\sum_{i=1}^n \hat{e}_i = 0$.

Demeaned regressors

- Sometimes it is useful to separate the constant from the other regressors: $Y_i = \boldsymbol{X}_i'\boldsymbol{\beta} + \alpha + e_i$, 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 $\sum_{i=1}^n \hat{e}_i = 0$ and $\sum_{i=1}^n \boldsymbol{X}_i\hat{e}_i = \mathbf{0}$ can be written as

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

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

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

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

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

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

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

  $\boldsymbol{Y}$ 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

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

- 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

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

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

Projection matrix

- 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.
- 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 $\boldsymbol{P}\boldsymbol{X}_1 = \boldsymbol{X}_1$.

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

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

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

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

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

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

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

- 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

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.

Orthogonal projection

- Define

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

- Note

$$\boldsymbol{M}\boldsymbol{X} = \left(\boldsymbol{I}_n - \boldsymbol{P}\right)\boldsymbol{X} = \boldsymbol{X} - \boldsymbol{P}\boldsymbol{X} = \boldsymbol{X} - \boldsymbol{X} = \boldsymbol{0}.$$

  Thus $\boldsymbol{M}$ and $\boldsymbol{X}$ are orthogonal. We call $\boldsymbol{M}$ the **orthogonal projection matrix**.

- If $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{\Gamma}$, then

$$\boldsymbol{M}\boldsymbol{Z} = \boldsymbol{Z} - \boldsymbol{P}\boldsymbol{Z} = \boldsymbol{0}.$$

  For example, $\boldsymbol{M}\boldsymbol{X}_1 = \boldsymbol{0}$ for any subcomponent $\boldsymbol{X}_1$ of $\boldsymbol{X}$ and $\boldsymbol{M}\boldsymbol{P} = \boldsymbol{0}$.

- $\boldsymbol{M}$ is symmetric ($\boldsymbol{M}' = \boldsymbol{M}$) and idempotent ($\boldsymbol{M}\boldsymbol{M} = \boldsymbol{M}$). $\mathrm{tr}\boldsymbol{M} = n - k$. The rank of $\boldsymbol{M}$ is $n - k$.

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

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

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

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

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

- We find

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

Estimation of error variance

- If e_i were observed, we would estimate $\sigma^2 = \mathbb{E}\left(e_i^2\right)$ by $\tilde{\sigma}^2 = n^{-1}\sum_{i=1}^n e_i^2$. This is infeasible as e_i is not observed.
- The feasible estimator: $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n \hat{e}_i^2$. In matrix notation, $\tilde{\sigma}^2 = n^{-1}\boldsymbol{e}'\boldsymbol{e}$ and $\hat{\sigma}^2 = n^{-1}\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}}$.
- Since $\hat{\boldsymbol{e}} = \boldsymbol{M}\boldsymbol{Y} = \boldsymbol{M}\boldsymbol{e}$,

$$\hat{\sigma}^2 = n^{-1}\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}} = n^{-1}\boldsymbol{Y}'\boldsymbol{M}\boldsymbol{M}\boldsymbol{Y} = n^{-1}\boldsymbol{Y}'\boldsymbol{M}\boldsymbol{Y} = n^{-1}\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e}.$$

- An implication:

$$\tilde{\sigma}^2 - \hat{\sigma}^2 = n^{-1}\boldsymbol{e}'\boldsymbol{e} - n^{-1}\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e} = n^{-1}\boldsymbol{e}'\boldsymbol{P}\boldsymbol{e} \geq 0.$$

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

$$\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} \hat{Y}_i^2 + \sum_{i=1}^{n} \hat{e}_i^2.$$

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

$$\boldsymbol{Y} - \mathbf{1}\bar{Y} = \hat{\boldsymbol{Y}} - \mathbf{1}\bar{Y} + \hat{\boldsymbol{e}},$$

where

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

- 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.

- A commonly reported statistic is the R^2:

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

  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.

Regression components

- Partition

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 \end{bmatrix}$$

and

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}.$$

- 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

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

- We show that

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

Residual regression

- Note

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

  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 $\boldsymbol{X}_1$.

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

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

$$\boldsymbol{M}_1 \left(\boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_2\hat{\boldsymbol{\beta}}_2 + \hat{\boldsymbol{e}}\right) - \boldsymbol{M}_1\boldsymbol{X}_2\hat{\boldsymbol{\beta}}_2 = \boldsymbol{M}_1\hat{\boldsymbol{e}} = \hat{\boldsymbol{e}}.$$