Finite-Sample Properties of OLS

Laura Malkhasyan, Aiwei Huang, Madhurima Chandra

University of Bonn

Outline

1. Lemma 3.1.1:
Orthogonal Projection matrices

2. Proposition 3.1.2:
OLS residuals

3. Proposition 3.1.3:
Variance Decomposition

Lemma 3.1.1: Orthogonal Projection matrices

Projection matrices

- The (OLS) fitted value: $\hat{y}_i = x_i b$
  In matrix notation: $\hat{y} = Xb \to \hat{y} = X(X'X)^{-1}X'y = Py$
  Where the (OLS) estimator: $b = (X'X)^{-1}X'y$

- The (OLS) residual: $\hat{\varepsilon}_i = y_i - \hat{y}_i$
  In matrix notation: $\hat{\varepsilon} = y - \hat{y} = \left(I_n - X(X'X)^{-1}X'\right) y = My$

Projection matrices

$\mathsf{P} = \mathsf{X}(\mathsf{X}'\mathsf{X})^{-1}\mathsf{X}'$ is called orthogonal projection matrix → projects any vector into the column space spanned by X

$\mathsf{M} = \mathsf{I}_n - \mathsf{X}(\mathsf{X}'\mathsf{X})^{-1}\mathsf{X}'$ is the associated orthogonal projection matrix → projects any vector into space orthogonal to span of X

Properties of projection matrices

Lemma 3.1.1 (Orthogonal Projection matrices) For $\mathsf{P} = \mathsf{X}(\mathsf{X}'\mathsf{X})^{-1}\mathsf{X}'$ and $\mathsf{M} = \mathsf{I}_n - \mathsf{P}$, where X is of full rank:

1. P and M are symmetric and idempotent:
    - $\mathsf{P} = \mathsf{P}'$ and $\mathsf{M} = \mathsf{M}'$
    - $\mathsf{P}\mathsf{P} = \mathsf{P} \quad \text{and} \quad \mathsf{M}\mathsf{M} = \mathsf{M}$

2. $\mathsf{X}'\mathsf{P} = \mathsf{X}', \quad \mathsf{X}'\mathsf{M} = \mathbf{0}, \quad \text{and} \quad \mathsf{P}\mathsf{M} = \mathbf{0}$

Proof

Can show these properties directly from the definitions of P and M.

1. - $\mathrm{P}' = (\mathrm{X}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}')' = \mathrm{X}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}' = \mathrm{P}$
     $\mathrm{M}' = \mathrm{I}_n - \mathrm{P}' = \mathrm{M}$
   - $\mathrm{PP} = \mathrm{X}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{X}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}' = \mathrm{X}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}' = \mathrm{P}$
     $\mathrm{MM} = (\mathrm{I}_n - \mathrm{P})(\mathrm{I}_n - \mathrm{P}) = \mathrm{I}_n - \mathrm{P} = \mathrm{M}$

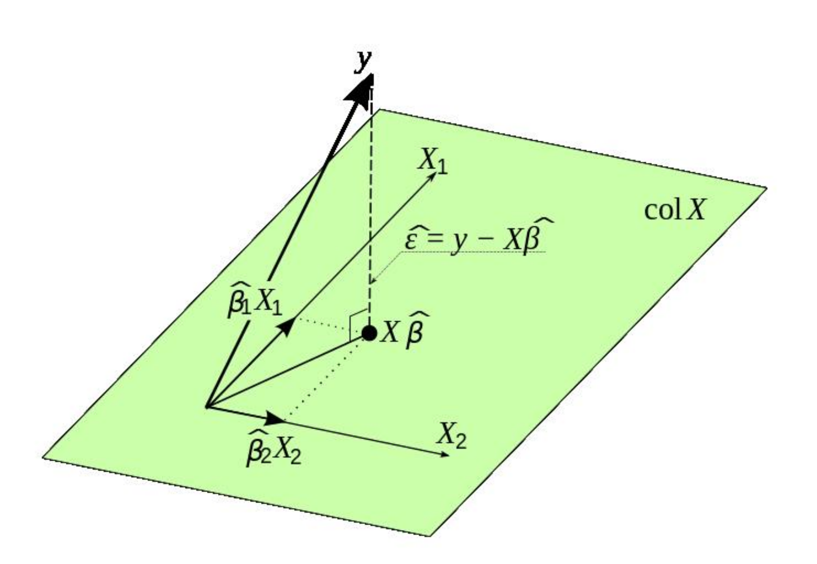
2. - $\mathrm{X}'\mathrm{P} = \mathrm{X}'\mathrm{X}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}' = \mathrm{X}'$
   - $\mathrm{X}'\mathrm{M} = \mathrm{X}'(\mathrm{I}_n - \mathrm{P}) = \mathrm{X}' - \mathrm{X}'\mathrm{P} = 0$
   - $\mathrm{PM} = \mathrm{P}(\mathrm{I}_n - \mathrm{P}) = 0$

Proposition 3.1.2: OLS residuals

OLS residuals

The n-vector $y - X\beta$ is called residual, with β allowed to vary arbitrarily.
The **OLS residual** is the one that orthogonal to the column space spanned by X or the subspace $\mathcal{S}(X_1, X_2, \ldots, X_K)$.

OLS residuals



OLS geometric interpretation.

Proposition 3.1.2 (OLS residuals)
For the OLS residuals and the OLS fitted values it holds that

$$\mathbf{X}'\hat{\varepsilon} = \mathbf{0}, \quad \text{and}$$
$$\mathbf{y}'\mathbf{y} = \hat{\mathbf{y}}'\hat{\mathbf{y}} + \hat{\varepsilon}'\hat{\varepsilon}.$$

Theorem and Proof

Proof.

$$\begin{aligned} X'\hat{\varepsilon} &= X'My \quad \text{(By Def. of } M\text{)} \\ &= 0y \quad \text{(By Lemma 3.1.1 part (ii))} \\ &= \underset{(K\times 1)}{0} \end{aligned}$$

$$\begin{aligned} y'y &= (Py + My)'(Py + My) \quad \text{(By Def. of } P \text{ and } M\text{)} \\ &= (y'P' + y'M')(Py + My) \\ &= y'P'Py + y'M'My + 0 \quad \text{(By Lemma 3.1.1 part (ii))} \\ &= \hat{y}'\hat{y} + \hat{\varepsilon}'\hat{\varepsilon} \end{aligned}$$

Unbiased Variance Estimator

Def.

$$s^2 = \frac{1}{n-K}\sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

Why $n-K$ degrees of freedom ?
$\hat{\varepsilon}_i^2$ loses K degrees of freedom because it has to satisfy the K linear restrictions ($\mathbf{X}'\hat{\boldsymbol{\varepsilon}} = \mathbf{0}$).

Proposition 3.1.3: Variance Decomposition

Variance Decomposition

Total sample variance of dependent variable (for a linear model with intercept) can be decomposed into variance explained by the regressors and variance explained by other factors unaccounted for in the model:

$$\underbrace{\sum_{i=1}^{n} (y_i - \bar{y})^2}_{\text{total variance}} = \underbrace{\sum_{i=1}^{n} \left(\hat{y}_i - \bar{\hat{y}}\right)^2}_{\text{explained variance}} + \underbrace{\sum_{i=1}^{n} \hat{\varepsilon}_i^{\,2}}_{\text{unexplained variance}}$$

Variance Decomposition: Proof

From Proposition 3.1.2, we know that $\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$, for regressions with intercept.
Hence, from $y_i = \hat{y}_i + \hat{\varepsilon}_i$ it follows that

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} y_i &= \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i + \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i \\ \bar{y} &= \bar{\hat{y}}_i + 0 \end{aligned}$$

Variance Decomposition: Proof (Contd.)

From Proposition 3.1.2, we know:

$$
\begin{aligned}
\mathbf{y}'\mathbf{y} &= \hat{\mathbf{y}}'\hat{\mathbf{y}} + \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} \\
\mathbf{y}'\mathbf{y} - n\bar{y}^2 &= \hat{\mathbf{y}}'\hat{\mathbf{y}} - n\bar{y}^2 + \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} \\
\mathbf{y}'\mathbf{y} - n\bar{y}^2 &= \hat{\mathbf{y}}'\hat{\mathbf{y}} - n\bar{\hat{y}}^2 + \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} \quad (\text{since } \bar{y} = \bar{\hat{y}}_i) \\
\sum_{i=1}^{n} y_i^2 - n\bar{y}^2 &= \sum_{i=1}^{n} \hat{y}_i^2 - n\bar{\hat{y}}^2 + \sum_{i=1}^{n} \hat{\varepsilon}_i^2 \\
\sum_{i=1}^{n} (y_i - \bar{y})^2 &= \sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}})^2 + \sum_{i=1}^{n} \hat{\varepsilon}_i^2
\end{aligned}
$$

Coefficient of Determination R^2

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{\hat{y}}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} = 1 - \frac{\sum_{i=1}^{n}\hat{u}_i^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$$

- Ratio of 'explained' variance to the 'total' variance of dependent variable.
- Larger the proportion of explained variance, better is the fit of the model.
- $0 \leq R^2 \leq 1$.
- Disadvantage of using R^2 as a measure of goodness-of-fit.