Properties of OLS Estimators

Econometrics II

Douglas G. Steigerwald

UC Santa Barbara

Overview

Reference: B. Hansen Econometrics Chapter 4.1-4.7

- Conditional Distribution of OLSE
  - mean
  - variance
- Unconditional Distribution of OLSE
  - mean
- Optimality of Least Squares Estimators
- Conditional moments of residuals and standardized residuals

Mean and Variance of Estimators

- begin with simple setting

$$y_i = \mu + u_i$$

  - $\mathbb{E}(u_i) = 0$ by construction $(u_i = e_i)$
  - corresponds to a model with a single covariate $x_i = 1$

- OLS estimator is $\overline{y}$
- mean of $\overline{y}$

$$\mathbb{E}(\overline{y}) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(y_i) = \mu$$

  - required assumption: $\mathbb{E}(y_i) = \mu$
  - the sample mean is unbiased, and hence the OLSE, is unbiased

An estimator $\widehat{\theta}$ for θ is (mean) unbiased if $\mathbb{E}\left(\widehat{\theta}\right) = \theta$.

Variance: Sample Mean

- begin with

$$\begin{aligned} y_i - \mu &= u_i \\ \overline{y} - \mu &= \frac{1}{n} \sum_{i=1}^{n} u_i \end{aligned}$$

- variance of $\overline{y}$

$$\mathbb{E} \left(\overline{y} - \mu \right)^2 = \frac{1}{n^2} \sum_{i=1}^{n} Var \left(u_i \right) = \frac{1}{n} \sigma^2$$

  - required assumption: $Var \left(u_i \right) = \sigma^2$
  - required assumption: $Cov \left(u_i, u_j \right) = 0$

OLSE General Setting

- general setting, k explanatory variables

Assumption 1 (Linear Regression Model).
The observations (y_i, x_i) come from a random sample and satisfy the linear regression equation

$$\begin{aligned} y_i &= x_i^{\mathrm{T}}\beta + u_i \\ \mathbb{E}\left(u_i | x_i\right) &= 0. \end{aligned}$$

The variables have finite second moments
$\mathbb{E}\left(y_i^2\right) < \infty$
$\mathbb{E}\left\|x_i\right\| < \infty,$
and an invertible design matrix
$\mathbb{E}\left(x_i x_i^{\mathrm{T}}\right) > 0.$

Mean: OLSE Summation Notation (Student Annotation)

Mean: OLSE Matrix Notation (Student Annotation)

Mean: OLSE Matrix Notation - Decomposition

- insert $y = X\beta + u$ into the formula for $\widehat{\beta}$

$$\begin{aligned}\widehat{\beta} &= \left(X^{\mathrm{T}}X\right)^{-1} X^{\mathrm{T}}\left(X\beta + u\right)\\ &= \beta + \left(X^{\mathrm{T}}X\right)^{-1} X^{\mathrm{T}}u\end{aligned}$$

  - useful linear decomposition of the estimator into β and stochastic component, $\left(X^{\mathrm{T}}X\right)^{-1} X^{\mathrm{T}}u$
- again we calculate

$$\mathbb{E}\left(\widehat{\beta} - \beta | X\right) = \left(X^{\mathrm{T}}X\right)^{-1} X^{\mathrm{T}}\mathbb{E}\left(u|X\right)$$

  - $\mathbb{E}\left(u|X\right) = \begin{pmatrix} \vdots \\ \mathbb{E}\left(u_i|X\right) \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ \mathbb{E}\left(u_i|x_i\right) \\ \vdots \end{pmatrix} = \mathbf{0}$
    - ★ $\mathbb{E}\left(\widehat{\beta} - \beta | X\right) = 0$

Unbiasedness of the OLS Estimator

- applying the law of iterated expectations

$$\mathbb{E}\left(\widehat{\beta}\right) = \mathbb{E}\left(\mathbb{E}\left(\widehat{\beta}|X\right)\right) = \beta$$

- we have shown the following theorem

Theorem (Mean of OLSE).

In the linear regression model of Assumption 1:

1. $\mathbb{E}\left(\widehat{\beta}\right) = \beta$;
2. $\mathbb{E}\left(\widehat{\beta}|X\right) = \beta$.

1) says the distribution of $\widehat{\beta}$ is centered at β
2) a stonger result, $\widehat{\beta}$ is unbiased for any realization of X

Variance and Conditional Variance Matrices

- for Z an $r \times 1$ random vector, define the $r \times r$ covariance matrix

$$\begin{aligned} Var\,(Z) &= \mathbb{E}\left((Z - \mathbb{E}Z)\,(Z - \mathbb{E}Z)^{\mathrm{T}}\right) \\ &= \mathbb{E}ZZ^{\mathrm{T}} - \mathbb{E}Z\,(\mathbb{E}Z)^{\mathrm{T}} \end{aligned}$$

- for any pair (Z, X) define the conditional covariance matrix

$$Var\,(Z|X) = \mathbb{E}\left((Z - \mathbb{E}\,(Z|X))\,(Z - \mathbb{E}\,(Z|X))^{\mathrm{T}}\,|X\right)$$

- conditional covariance matrix of the OLSE

$$V := Var\left(\widehat{\beta}|X\right)$$

Error Variance Assumptions

- we consider the general case of heteroskedastic regression

$$\mathbb{E}\left(u_i^2 | x_i\right) = \sigma^2\left(x_i\right) = \sigma_i^2$$

- we also consider the specialized case of homoskedastic regression

Assumption 2 (Homoskedastic Linear Regression Model). In addition to Assumption 1,

$$\mathbb{E}\left(u_i^2 | x_i\right) = \sigma^2\left(x_i\right) = \sigma^2,$$

is independent of x_i.

Conditional Covariance Matrix of the Error

- conditional covariance matrix of $n \times 1$ error u is the $n \times n$ matrix

$$D = \mathbb{E}\left(uu^{\mathrm{T}}|X\right)$$

  - *i'th* diagonal element $\mathbb{E}\left(u_i^2|X\right) = \mathbb{E}\left(u_i^2|x_i\right) = \sigma_i^2$
  - *ij'th* off-diagonal element $\mathbb{E}\left(u_i u_j|X\right) = \mathbb{E}\left(u_i|x_i\right)\mathbb{E}\left(u_j|x_j\right) = 0$

- D is a diagonal matrix

$$D = diag(\sigma_1^2, \ldots, \sigma_n^2) = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

- under homoskedasticity

$$\begin{aligned} \mathbb{E}\left(u_i^2|x_i\right) &= \sigma^2\left(x_i\right) = \sigma^2 \\ D &= I_n\sigma^2 \end{aligned}$$

Conditional Variance of Linear Combinations

- for any $n \times r$ matrix $A = A(X)$

$$Var\left(A^{\mathrm{T}}y|X\right) = Var\left(A^{\mathrm{T}}u|X\right) = A^{\mathrm{T}}DA$$

- in particular $\widehat{\beta} = A^{\mathrm{T}}y$
  - $A = X\left(X^{\mathrm{T}}X\right)^{-1}$

$$V = Var\left(\widehat{\beta}|X\right) = A^{\mathrm{T}}DA = \left(X^{\mathrm{T}}X\right)^{-1}X^{\mathrm{T}}DX\left(X^{\mathrm{T}}X\right)^{-1}$$

  - sandwich form
  - $X^{\mathrm{T}}DX = \sum_{i=1}^{n} x_i^2\sigma_i^2$
    - weighted version of $X^{\mathrm{T}}X$

- under homoskedasticity, $D = I_n\sigma^2$

$$\begin{aligned} X^{\mathrm{T}}DX &= X^{\mathrm{T}}X\sigma^2 \\ V &= \left(X^{\mathrm{T}}X\right)^{-1}\sigma^2 \end{aligned}$$

Variance of OLSE

Theorem (Variance of the OLSE).

In the linear regression model of Assumption 1:

$$V := Var\left(\widehat{\beta}|X\right) = \left(X^{\mathrm{T}}X\right)^{-1} X^{\mathrm{T}}DX \left(X^{\mathrm{T}}X\right)^{-1}$$

with $D = diag(\sigma_1^2, \ldots, \sigma_n^2)$.

In the homoskedastic linear regression model of Assumption 2
$V = \left(X^{\mathrm{T}}X\right)^{-1}\sigma^2$.

Gauss-Markov Theorem: Linear Unbiased Estimators

- consider the class of estimators of β that are linear functions of y

$$\widetilde{\beta} = A^{\mathrm{T}} y$$

  - A an $n \times k$ function of X
  - OLS is a special case with $A = \left(X^{\mathrm{T}} X\right)^{-1} X$
- what is the best choice of A?
  - best means smallest variance
- Gauss-Markov Theorem: $A = \left(X^{\mathrm{T}} X\right)^{-1} X$ is the best choice for unbiased, homoskedastic models

Theorem (Gauss-Markov).

1. In the linear regression model of Assumption 1:
The best linear unbiased estimator is

$$\widetilde{\beta} = \left(X^{\mathrm{T}} D^{-1} X\right)^{-1} X^{\mathrm{T}} D^{-1} y.$$

2. In the homoskedastic linear regression model of Assumption 2:
The best linear unbiased estimator is the OLSE

$$\widehat{\beta} = \left(X^{\mathrm{T}} X\right)^{-1} X^{\mathrm{T}} y.$$

Gauss-Markov Theorem Interpretation

1. for linear (heteroskedastic) regression models
    1. BLUE is $\widetilde{\beta}$ - the Generalized LSE (not OLSE)
    2. infeasible as D is unknown
    3. need an estimator of D to have a practical alternative to $\widehat{\beta}$
2. for linear homoskedastic regression models
    1. BLUE is $\widehat{\beta}$
        1. special case of $\widetilde{\beta}$ with $D = I_n \sigma^2$
    2. limited efficiency result
        1. restricted to homoskedastic models and linear unbiased estimators
        2. could be biased or nonlinear estimators with lower MSE

Proof of Gauss-Markov Theorem

Residuals

- residuals

$$\begin{aligned} \widehat{u}_i &= y_i - x_i^{\mathrm{T}} \widehat{\beta} \\ \widehat{u} &= My = Mu \end{aligned}$$

  - $M = I_n - X \left(X^{\mathrm{T}} X \right)^{-1} X^{\mathrm{T}}$ and $MX = \mathbf{0}$

- conditional mean

$$\mathbb{E}\left(\widehat{u}|X\right) = \mathbb{E}\left(Mu|X\right) = M\mathbb{E}\left(u|X\right) = 0$$

- conditional variance

$$Var\left(\widehat{u}|X\right) = M \cdot Var\left(u|X\right) \cdot M = MDM$$

Conditional Variance of Residuals

- under conditional homoskedasticity (Assumption 2)

$$\begin{aligned} \mathbb{E}\left(u_i^2|x_i\right) &= \sigma^2 \\ Var\left(\widehat{u}|X\right) &= M\sigma^2 \end{aligned}$$

  - this follows from the fact that M is idempotent (and symmetric)
- u is homoskedastic but $\widehat{u}$ is heteroskedastic
  - conditional variance equals M not $I_n\sigma^2$
- i'th diagonal element of $M\sigma^2$ is $Var\left(\widehat{u}_i|X\right)$

$$Var\left(\widehat{u}_i|X\right) = \mathbb{E}\left(\widehat{u}_i^2|X\right) = \left(1 - h_{ii}\right)\sigma^2$$

  - $h_{ii} = x_i^{\mathrm{T}}\left(X^{\mathrm{T}}X\right)^{-1}x_i$
    - ★ a function of x_i, therefore residuals are heteroskedastic even if errors are homoskedastic

Standardized Residuals

- rescale to get constant conditional variance

$$\overline{u}_i = (1 - h_{ii})^{-1/2} \, \widehat{u}_i$$

$$Var(\overline{u}_i | X) = \sigma^2$$

- standardized residuals have conditional mean and conditional variance that are identical to the conditional mean and conditional variance of the errors u

Error Variance Estimator

- error variance $\sigma^2 := \mathbb{E}\left(u_i^2\right)$
- method-of-moments estimator

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\widehat{u}_i^2$$

  - MLE for normal regression model

$$\widehat{\sigma}^2 = \frac{1}{n}\widehat{u}^{\mathrm{T}}\widehat{u} = \frac{1}{n}u^{\mathrm{T}}Mu = \frac{1}{n}tr\left(u^{\mathrm{T}}Mu\right) = \frac{1}{n}tr\left(Muu^{\mathrm{T}}\right)$$

- conditional expectation

$$\begin{aligned}\mathbb{E}\left(\widehat{\sigma}^2|X\right) &= \frac{1}{n}tr\left(\mathbb{E}\left(Muu^{\mathrm{T}}|X\right)\right)\\ &= \frac{1}{n}tr\left(M\mathbb{E}\left(uu^{\mathrm{T}}|X\right)\right)\\ &= \frac{1}{n}tr\left(MD\right)\end{aligned}$$

Error Variance Estimator Bias

- under conditional homoskedasticity $D = I_n \sigma^2$

$$\begin{aligned} \mathbb{E}\left(\widehat{\sigma}^2 | X\right) &= \frac{1}{n} tr\left(M\sigma^2\right) \\ &= \left(\frac{n-k}{n}\right)\sigma^2 \end{aligned}$$

- another way to see this

$$\begin{aligned} \mathbb{E}\left(\widehat{\sigma}^2 | X\right) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left(\widehat{u}_i^2 | X\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} \left(1 - h_{ii}\right)\sigma^2 \\ &= \left(\frac{n-k}{n}\right)\sigma^2 \end{aligned}$$

$\widehat{\sigma}^2$ is biased toward zero by a factor of $\frac{k}{n}$

Unbiased Error Variance Estimation

- bias takes scale form, therefore rescale to get unbiased estimator

$$s^2 = \frac{n}{n-k}\widehat{\sigma}^2 = \frac{1}{n-k}\sum_{i=1}^{n}\widehat{u}_i^2$$

  - by the earlier calculation $\mathbb{E}\left(s^2|X\right) = \sigma^2$ so $\mathbb{E}\left(s^2\right) = \sigma^2$
    - ★ bias-corrected estimator, widely used

- bias-corrected estimator can also be constructed from standardized residuals

$$\begin{aligned}\overline{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}\overline{u}_i^2 = \frac{1}{n}\sum_{i=1}^{n}\left(1-h_{ii}\right)^{-1}\widehat{u}_i^2 \\ \mathbb{E}\left(\overline{\sigma}^2|X\right) &= \sigma^2 \;\text{ so }\; \mathbb{E}\left(\overline{\sigma}^2\right) = \sigma^2\end{aligned}$$

- if n is not large relative to k, use a bias-corrected estimator

Proof of Gauss-Markov Theorem

- let A be any $n \times k$ function of X with $\widetilde{\beta} = A^{\mathrm{T}} y$
  - under Assumption 2 , $Var\left(A^{\mathrm{T}} y | X\right) = A^{\mathrm{T}} A \sigma^2$
  - $\widehat{\beta}$ is efficient if $\quad A^{\mathrm{T}} A - \left(X^{\mathrm{T}} X\right)^{-1}$ is a positive semi-definite matrix
  - let $C = A - X\left(X^{\mathrm{T}} X\right)^{-1}$
- $A^{\mathrm{T}} A - \left(X^{\mathrm{T}} X\right)^{-1}$ equals
- $= \left(C + X(X^{\mathrm{T}} X)^{-1}\right)^{\mathrm{T}} \left(C + X(X^{\mathrm{T}} X)^{-1}\right) - (X^{\mathrm{T}} X)^{-1}$
- $= C^{\mathrm{T}} C - C^{\mathrm{T}} X (X^{\mathrm{T}} X)^{-1} - (X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} C + (X^{\mathrm{T}} X)^{-1} - (X^{\mathrm{T}} X)^{-1}$
  - note $\mathbb{E}\left(\widetilde{\beta} | X\right) = A^{\mathrm{T}} X \beta + A^{\mathrm{T}} \mathbb{E}\left(u | X\right) = A^{\mathrm{T}} X \beta$
  - $\widetilde{\beta}$ unbiased $\Rightarrow A^{\mathrm{T}} X = I_k \Rightarrow C^{\mathrm{T}} X = 0$
- $A^{\mathrm{T}} A - \left(X^{\mathrm{T}} X\right)^{-1} = C^{\mathrm{T}} C$
  - if $M = C^{\mathrm{T}} C$ then M is positive semi-definite, as required

Return to Gauss-Markov Theorem