Algebra of Least Squares

Econometrics II

Douglas G. Steigerwald

UC Santa Barbara

Overview

Reference: B. Hansen Econometrics Chapter 3.1-3.9

- Introduce Observational Data
- Regression (Best Linear Projection) Estimation
- How do we estimate β?
- Define OLSE (method-of-moments estimator)
- Derive OLS Residual Sample Properties

Observational Data - Random Samples

- (observational) data
  - to emphasize data notationally, introduce subscript i

$$\{(y_i, x_i)\, ; i = 1, \ldots n\}$$

- data - realization of a random process
  - must begin with an assumption about the data generating process
- Assumption: $\{(y_1, x_1), \ldots, (y_n, x_n)\}$ are a random sample
  - independent and identically distributed (iid)
  - ordering is irrelevant
  - nothing special about any specific observation
  - rarely true for observational data in economics

Linear Projection Model with Data

- regression model applies to the observations

$$y_i = x_i^{\mathrm{T}}\beta + u_i$$

  - β_{lpc} is the value of β that minimizes $\mathbb{E}\left(y_i - x_i^{\mathrm{T}}\beta\right)^2$
  - explicit solution

$$\beta_{lpc} = \left(\mathbb{E}\left(x_i x_i^{\mathrm{T}}\right)\right)^{-1}\mathbb{E}\left(x_i y_i\right)$$

  - yielding the projection

$$\mathcal{P}\left(y_i | x_i\right) = x_i^{\mathrm{T}}\beta_{lpc}$$

Method of Moments Estimation

- example: unconditional mean of y_i

$$y_i = \mu + e_i$$

  - $\mathbb{E}\left(e_i\right) = 0$
  - $\mu = \mathbb{E}\left(y_i\right)$

- moment estimator $\widehat{\mu}$

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

- sample analog for population moment

$$\begin{aligned} \widehat{\mu} &= \sum_{i=1}^{n} y_i \frac{1}{n} \\ \mu &= \int y f_y\left(y\right) dy \end{aligned}$$

Least Squares Estimator

- β_{lpc} minimizes $S(\beta) = \mathbb{E}\left(y_i - x_i^{\mathrm{T}}\beta\right)^2$
- $\widehat{\beta}_{lpc}$ minimizes $S_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i^{\mathrm{T}}\beta\right)^2$
  - $\widehat{\beta}_{lpc}$ is a moment estimator
- define the sum-of-squared-errors function

$$SSE_n(\beta) := \sum_{i=1}^{n}\left(y_i - x_i^{\mathrm{T}}\beta\right)^2$$

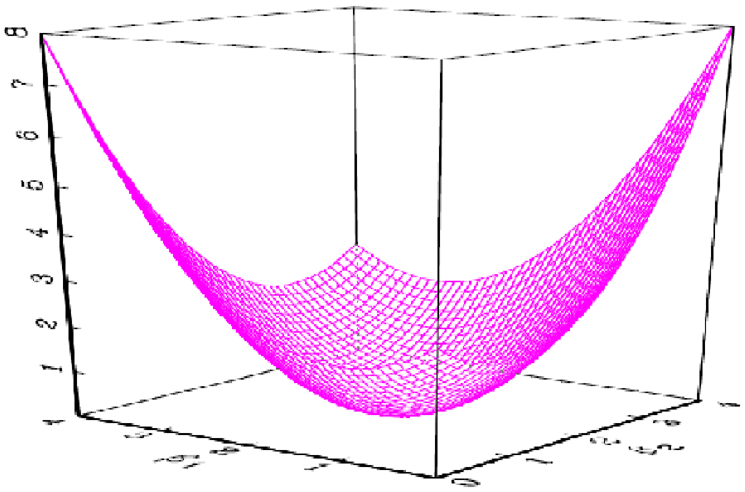
- because $S_n(\beta)$ is a scale multiple of $SSE_n(\beta)$

$$\widehat{\beta}_{lpc} = \arg\min_{\beta\in\mathbb{R}^k} SSE_n(\beta)$$

  - $\widehat{\beta}_{lpc}$ commonly called the ordinary least squares (OLS) estimator
    - ★ denoted $\widehat{\beta}_{ols}$ or simply $\widehat{\beta}$

Sum-of-Squared Errors Function

given $k = 2$ and a random sample



One Regressor (Student Annotation)

Multiple Regressors

- $SSE_n(\beta) = \sum_{i=1}^{n} y_i^2 - 2\beta^{\mathrm{T}} \sum_{i=1}^{n} x_i y_i + \beta^T \sum_{i=1}^{n} x_i x_i^{\mathrm{T}} \beta$
- FOC (k equations in k unknowns)

$$0 = \frac{\partial}{\partial \beta} SSE_n(\beta) = -2 \sum_{i=1}^{n} x_i y_i + 2 \sum_{i=1}^{n} x_i x_i^{\mathrm{T}} \widehat{\beta}$$

- matrix algebra yields compact solution

$$\widehat{\beta} = \left(\sum_{i=1}^{n} x_i x_i^{\mathrm{T}} \right)^{-1} \sum_{i=1}^{n} x_i y_i$$

- $\widehat{\beta}$ is a method-of-moments estimator, sample analog for

$$\beta = \left(\mathbb{E}\left(x_i x_i^{\mathrm{T}} \right) \right)^{-1} \mathbb{E}(x_i y_i)$$

Ordinary Least Squares Estimator

$$\widehat{\beta} = \underset{\beta \in \mathbb{R}^k}{\arg\min} S_n(\beta)$$

where $S_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i^{\mathrm{T}}\beta\right)^2$ has the solution

$$\widehat{\beta} = \left(\sum_{i=1}^{n} x_i x_i^{\mathrm{T}}\right)^{-1} \sum_{i=1}^{n} x_i y_i$$

- first published by Adrien-Marie Legendre (1805)
  - existing problem in astronomical measurement
    - solve system of n equations with $k < n$ unknowns (e.g. $y_i = x_i^{\mathrm{T}}\beta + u_i$)
  - noted FOC a system of k equations in k unknowns
  - could be solved by ordinary methods
  - hence ordinary least squares

Illustration: Bivariate Regression

- data from March 2009 Current Population Survey
  - sub-sample married (spouse present) black female wage earners
  - 12 years of potential work experience
    - 20 observations
- y_i - log wages x_i - years of education and intercept

$$\sum_{i=1}^{n} x_i y_i = \begin{pmatrix} 995.86 \\ 62.64 \end{pmatrix}$$

$$\sum_{i=1}^{n} x_i x_i^{\mathrm{T}} = \begin{pmatrix} 5010 & 314 \\ 314 & 20 \end{pmatrix}$$

Illustration: Solution

$$\left(\sum_{i=1}^{n} x_i x_i^{\mathrm{T}}\right)^{-1} = \begin{pmatrix} 0.0125 & -0.196 \\ -0.196 & 3.124 \end{pmatrix}$$

$$\begin{aligned} \widehat{\beta} &= \begin{pmatrix} 0.0125 & -0.196 \\ -0.196 & 3.124 \end{pmatrix} \begin{pmatrix} 995.86 \\ 62.64 \end{pmatrix} \\ &= \begin{pmatrix} 0.155 \\ 0.698 \end{pmatrix} \end{aligned}$$

- display as

$$\widehat{\log(wage)} = 0.155 \; education + 0.698$$

  - each year of education associated with 16% increase in mean wage

Illustration: Multivariate Regression

- include all levels of experience
  - sub-sample: single (never married) Asian male wage earners
    - 268 observations
- covariates
  - years of education
  - years of potential work experience - *experience*
  - its square - $experience^2/100$
    - divide by 100 to simplify reporting
- estimated results
- $=$0.143 education $+$ 0.036 experience - 0.071 experience^2/100 $+$ 0.575
  - 14% increase in mean wages per year of education, holding experience constant

OLS Residuals

- residual

$$\widehat{u}_i = y_i - x_i^{\mathrm{T}}\widehat{\beta} := y_i - \widehat{y}_i$$

  - $\widehat{y}_i$ - fitted value, *not* predicted value
    - a function of the entire sample, including y_i, so not a valid prediction

- distinguish the regression error from the residual
  - error u_i is latent

- residual properties
  - FOC for OLSE

$$\sum_{i=1}^{n} x_i \left(y_i - x_i^{\mathrm{T}}\widehat{\beta} \right) = 0 \rightarrow \sum_{i=1}^{n} x_i \widehat{u}_i = 0$$

    - when x_i contains a constant, $\sum_{i=1}^{n} \widehat{u}_i = 0$

  - sample correlation between covariates and residuals is zero
  - sample mean of residuals is zero

- these are algebraic results and hold for all linear regression estimates

Model in Matrix Notation

- there are n equations, one for each observation

$$\begin{aligned} y_1 &= x_1^{\mathrm{T}}\beta + u_1 \\ y_2 &= x_2^{\mathrm{T}}\beta + u_2 \\ &\vdots \\ y_n &= x_n^{\mathrm{T}}\beta + u_n \end{aligned}$$

- conveniently written as

$$y = X\beta + u$$

  - $\underset{n\times 1}{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \underset{n\times k}{X} = \begin{pmatrix} x_1^{\mathrm{T}} \\ \vdots \\ x_n^{\mathrm{T}} \end{pmatrix} \quad \underset{n\times 1}{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$

Sample Sums in Matrix Notation

- $\sum_{i=1}^{n} x_i x_i^{\mathrm{T}} = X^{\mathrm{T}} X$
- $\sum_{i=1}^{n} x_i y_i = X^{\mathrm{T}} y$

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

- $\widehat{u} = y - X\widehat{\beta}$
- $X^{\mathrm{T}} u = 0$
- first known treatment of use of matrix methods to solve simultaneous system - *The Nine Chapters on the Mathematical Art* - Chapter 8
  - Chinese text, written by several generations of scholars 10th to 2nd century BCE

Review (Student Annotation)

- What is the regression model for observed data?

What function does the OLS estimator minimize?
What is the sample mean of the residuals?
What is the sample covariance between the residuals and the covariates?
How long ago were matrix methods used to solve systems of equations?
at least the 2nd century BCE and perhaps the 10th century BCE