Lecture 1

Matthew Osborne

5 September 2006

An Introduction to OLS

- Isolate the impact of changing some variable on another variable of economic interest, holding everything else fixed.
  - E.g. Race or gender on earnings.

- Problem: most data used in empirical economics is non experimental.
  - Survey data (CPS, NLSY).
  - Price, quantity data.
  - GDP data.

An Introduction to OLS

- Simple example: Effect of gender on wages (y).

- Sample of N people. N_m males, N_f females.

- Impact of being female on wages?

$$\overline{y}_f - \overline{y}_m = \frac{1}{N_f}\sum_{n=1}^{N_f} y_n - \frac{1}{N_m}\sum_{n=1}^{N_m} y_n$$

- Problem: gender may not be the only variable that affects wages.

An Introduction to OLS

- Suppose more men are union members than women, more men are white than women.

- These will also affect wages, and our previous strategy won't work.

- *Ordinary Least Squares* allows us to separate difference in wages among characteristics.

An Introduction to OLS

- Assume that we can describe the relationship between wages and attributes as:

$$y = x_1\beta_1 + x_2\beta_2 + ... + x_K\beta_K + \varepsilon$$

- y: LHS variables, x's: RHS variables.

- β's: coefficients.

- ε: residual or error term.

- OLS finds the β's that solve the following:

$$\min_{\beta_1,\ldots,\beta_K} \sum_{n=1}^{N} [y_n - (x_1\beta_1 + x_2\beta_2 + ... + x_K\beta_K)]^2$$

Simple Example

- Suppose $K = 1$, $x_{n1} = 1$.

$$\min_{\beta_1} \sum_{n=1}^{N} (y_n - \beta_1)^2$$

- Solution:

$$\widehat{\beta}_1 \equiv \operatorname*{argmin}_{\beta_1} \sum_{n=1}^{N} (y_n - \beta_1)^2 = \frac{\sum_{n=1}^{N} y_n}{N}$$

Simple Example 2

- Suppose $K = 2$, $x_{n1} = 1$,

$$x_{n2} = \begin{cases} 0 & \text{if person } n \text{ is male.} \\ 1 & \text{if person } n \text{ is female.} \end{cases}$$

$$\min_{\beta_1,\beta_2} \sum_{n=1}^{N} (y_n - \beta_1 - \beta_2 x_{n2})^2$$

$$= \min_{\beta_1,\beta_2} \left[\sum_{\{n|x_{n2}=0\}} (y_n - \beta_1)^2 + \sum_{\{n|x_{n2}=1\}} (y_n - \beta_1 - \beta_2)^2 \right]$$

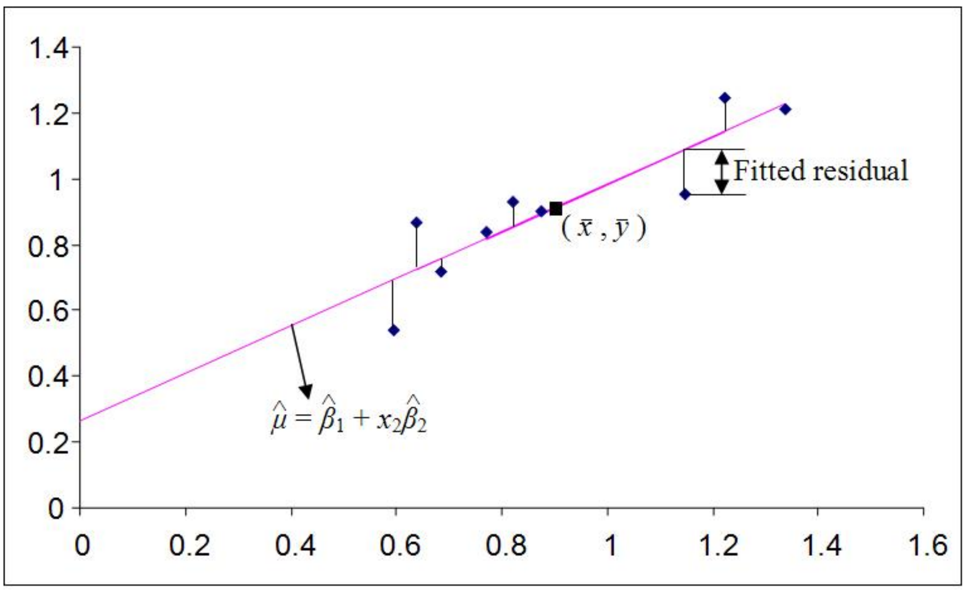
$$= \left[\min_{\beta_1} \sum_{\{n|x_{n2}=0\}} (y_n - \beta_1)^2 \right] + \left[\min_{\gamma} \sum_{\{n|x_{n2}=1\}} (y_n - \gamma)^2 \right]$$

- $\widehat{\beta}_2$ is the difference between average wages.

Generalized Sample Average

- OLS fitted equation:

$$\widehat{\mu} = x_1\widehat{\beta}_1 + x_2\widehat{\beta}_2 + ... + x_K\widehat{\beta}_K$$



The OLS Fit

- Adding variables decreases SSR:

$$\min_{\{\beta_1,\ldots,\beta_K\}} \sum_{n=1}^{N} \left[y_n - \left(\sum_{k=1}^{K} x_{nk}\beta_k \right) \right]$$
$$\geq \min_{\{\beta_1,\ldots,\beta_{K+1}\}} \sum_{n=1}^{N} \left[y_n - \left(\sum_{k=1}^{K+1} x_{nk}\beta_k \right) \right]$$

- A measure of fit, $R^2 = \widehat{\text{corr}}(y, \widehat{\mu})$.

$$R^2 = 1 - \frac{\sum_{n=1}^{N}(y_n - \widehat{\mu}_n)^2}{\sum_{n=1}^{N}(y_n - \overline{y}_n)^2}$$

OLS in Matrix Form

$$\sum_{k=1}^{K} x_{nk}\widehat{\beta}_k = \left[\begin{array}{ccc} x_{n1} & \cdots & x_{nK} \end{array}\right] \left[\begin{array}{c} \widehat{\beta}_1 \\ \vdots \\ \widehat{\beta}_K \end{array}\right] = \mathbf{x}_n'\hat{\boldsymbol{\beta}} \equiv \widehat{\mu}_n$$

- Linear in the β's.

- Column vector of LHS: **y**.

$$\mathbf{X} = \left[\begin{array}{c} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_N' \end{array}\right] = \left[\begin{array}{cccc} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NK} \end{array}\right] = [x_{nk}]$$

OLS in Matrix Form

- Column vector of ones: $\boldsymbol{\iota}_N$.
- $N \times N$ identity matrix: $\mathbf{I}_n$.

$$\overline{y} = \frac{\boldsymbol{\iota}'\mathbf{y}}{\boldsymbol{\iota}'\boldsymbol{\iota}}$$

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