Lab 1: Ordinary Least Squares

Econometrics Beyond Ordinary Least Squares

Patrick Mokre

WS 2020/2021

OLS: Intuition

profits = prices * output - wages * labor force - interest * fixed capital.

You can observe gross operating surplus GOS (profits), gross output GO (prices * unit output), total fixed assets (fixed capital) and total employee compensation (wages * labor force) from **national accounts**.

GOS = (**GO** - **TCE**) - interest rate * **FA**

Now the interest rate is the solution to one equation with one unknown.

$$i = \frac{GOS - GO + TCE}{FA} \tag{1}$$

Call GOS the **dependent variable**, (GO - TCE) an **intercept**, FA the **covariate** and i a **coefficient**.

More than one, otherwise identical, observation $\Rightarrow$ average interest rate on fixed capital.

OLS: Mathematics

Let $\tilde{y}$ be a weighted linear combination of some factors $x_2, ..., x_k$ (omit x_1 for a constant intercept).

Choose variable weigths $\tilde{\beta}_1, ...\tilde{\beta}_k$.

$$\tilde{y} = \tilde{\beta}_1 + \tilde{\beta}_2 x_2 + ... + \tilde{\beta}_k x_k \tag{2}$$

Deviation between y and $\tilde{y}$:

$$y - \tilde{y} = y - [\tilde{\beta}_1 + \tilde{\beta}_2 x_2 + ... + \tilde{\beta}_k x_k] \tag{3}$$

Vector Notation

Vector notation allows to express some amount of 1×1 values (eg. numbers) under one name.

Vectors have a **dimensionality** of rows $\times$ columns. One of the two is supposed to be 1.

Transposing a row vector makes it a column vector and vice versa.

$$\begin{bmatrix} z_1 & z_2 & z_3 \end{bmatrix}' = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}$$

OLS: Squared Loss Function

Deviation with vector notation, x and $\tilde{\beta}$ are $K \times 1$ vectors.

$$y - \tilde{y} = y - x'\tilde{\beta} \tag{4}$$

Introduce N combinations of y and x and denote each with index $i \in N$.

$$\tilde{y}_i = \tilde{\beta}_1 + \tilde{\beta}_2 x_{i,2} + ... + \tilde{\beta}_k x_{i,k} \tag{5}$$

Squared Loss function:

$$S(\tilde{\beta}) = \sum_{i=1}^{N} (y_i - x_i'\tilde{\beta})^2 \tag{6}$$

OLS: Minimization Problem

Find the **global minimum** of $S(\tilde{\beta})$ to determine the **best-fit** coefficient vector $\hat{\beta}$.

Minimize squared loss function $\Rightarrow$ Ordinary Least Squares.

$$\frac{\partial S(\tilde{\beta})}{\partial \tilde{\beta}} = -2 \sum x_i (y_i - x_i' \tilde{\beta}) = 0 \tag{7}$$

$$\left(\sum x_i x_i' \right) \tilde{\beta} = \left(\sum x_i y_i \right) \tag{8}$$

$$\hat{\beta} = \frac{\left(\sum x_i y_i \right)}{\left(\sum x_i x_i' \right)} = \left(\sum x_i x_i' \right)^{-1} \left(\sum x_i y_i \right) \tag{9}$$

Then the **best-fit** estimation for $\tilde{y}$ is given by vectors x and $\hat{\beta}$.

$$\hat{y} = x_i' \hat{\beta} \tag{10}$$

Matrix Notation

A matrix Z can be imagined as a **stacking** M row vectors with dimension $1 \times N$:
A $M \times N$ matrix

$$z_1 = \begin{bmatrix} z_{1,1} & z_{1,2} & z_{1,3} \end{bmatrix}$$
$$z_2 = \begin{bmatrix} z_{2,1} & z_{2,2} & z_{2,3} \end{bmatrix}$$

$$Z = \begin{bmatrix} z_{1,1} & z_{1,2} & z_{1,3} \\ z_{2,1} & z_{2,2} & z_{2,3} \end{bmatrix}$$

OLS: Matrix Result

$$\hat{\beta} = (X'X)^{-1}X'y \tag{11}$$

To derive $\hat{\beta}$, on must **invert** X. X must be **invertible**. X is invertible if no column is a linear combination of another, "**no multi-collinearity**".

$\Rightarrow$ Do not include the same covariates twice.

OLS: Residual

Residual $e_i = y_i - \hat{y}_i = y_i - x_i'\hat{\beta}$.

$$S(\hat{\beta}) = \sum_{i=1}^{N} e_i^2 \tag{12}$$

The $N \times 1$ vector e and $N \times K$ vector x are **orthogonal**.

$$\sum x_i(y_i - x_i'\hat{\beta}) = \sum x_i e_i = 0 \tag{13}$$

This means, the **average residual** is zero. If it wasn't, the approximation would not be ideal.

This also means that the linear approximation for y holds in the average.

$$\bar{y} = \bar{x}'\hat{\beta} \tag{14}$$

OLS: Simple Linear Regression

$$\tilde{\beta} = \tilde{\beta}_1, \tilde{\beta}_2 \tag{15}$$

$$y_i = \tilde{\beta}_1 + \tilde{\beta}_2 x_{i,2} \tag{16}$$

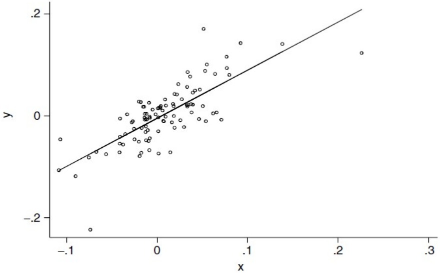


Figure 1: Verbeek, 2004, Figure 2.1: "Simple linear regression: fitted line and observation points"

$$S(\tilde{\beta}_1, \tilde{\beta}_2) = \sum_{i=1}^{N} (y_i - \tilde{\beta}_1 - x_i' \tilde{\beta}_2)^2 \tag{17}$$

OLS: Simple Linear Regression, Analytical Solution

$$S(\tilde{\beta}_1, \tilde{\beta}_2) = \sum_{i=1}^{N} (y_i - \tilde{\beta}_1 - x_i' \tilde{\beta}_2)^2 \tag{18}$$

$$\frac{\partial S}{\partial \beta_1} = -2 \sum (y_i - \tilde{\beta}_1 - \tilde{\beta}_2 x_i)^2 = 0 \tag{19}$$

$$\frac{\partial S}{\partial \beta_2} = -2 \sum x_i (y_i - \tilde{\beta}_1 - \tilde{\beta}_2 x_i)^2 = 0 \tag{20}$$

Analytical Solution for estimators:

$$\hat{\beta}_1 = \frac{1}{N} \sum y_i - \hat{\beta}_2 \frac{1}{N} \sum x_i = \bar{y} - \hat{\beta}_2 x_i \tag{21}$$

$$\sum x_i y_i - \hat{\beta}_1 \sum x_i - \hat{\beta}_2 \left(\sum x_i^2 \right) = 0 \tag{22}$$

$$\sum x_i y_i - N \bar{x} \bar{y} - \hat{\beta}_2 \left(\sum x_i^2 - N \bar{x}^2 \right) = 0 \tag{23}$$

$$\hat{\beta}_2 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \tag{24}$$

OLS: Practical Example

i	x	y
1	1.33	1.99
2	1.86	2.79
3	2.86	4.30
4	4.54	6.81
5	1.01	1.51

Please:

- Produce a plot with x on the x-axis and y on the y-axis
- Calculate $\bar{y}$, $\bar{x}$, $\hat{\beta}_1$ and $\hat{\beta}_2$.
- Plot the regression line and provide **graphical interpretations** of $\hat{\beta}_1$ and $\hat{\beta}_2$.
- Calculate the residual vector e.

OLS: Statistics

An **algebraic model** $\tilde{y}_i = \tilde{\beta}_1 + \tilde{\beta}_2 x_{i,2} + ... + \tilde{\beta}_k x_{i,k}$ is valid for and only for $i \in N$.

A **statistical model** is the attempt to describe a relationship that holds for all members of a well-defined set, eg. all German households.

$$y_i = \beta_1 + \beta_2 x_{i,2} + ... + \beta_k x_{i,k} + \epsilon_i \tag{25}$$

$$y_i = x_i' \beta + \epsilon_i \tag{26}$$

$$y = X\beta + \epsilon \tag{27}$$

The main difference is the unobserved **error term** ϵ_i.

Are such models **meaningful**?

OLS: Error Term Assumptions

Assume **exogenous covariates**: $E(y_i \mid x_i) = x_i'\beta \Rightarrow E(\epsilon_i \mid x_i) = 0$.

Now, β_k has a meaning: It is the marginal *ceteris paribus* effect of a change in $x_{i,k}$ on y_i, or more generally the marginal effect of a change in $\bar{x_k}$ on $\bar{y}$.

$\hat{\beta} = (X'X)^{-1}X'y$ is called the **ordinary least squares estimator** for β.

OLS: Statistical Example

i	x	y
1	1.33	2.12
2	1.86	2.83
3	2.86	4.14
4	4.54	6.72
5	1.01	1.48

- Produce a plot with x on the x-axis and y on the y-axis
- Calculate $\bar{y}$, $\bar{x}$, $\hat{\beta}_1$ and $\hat{\beta}_2$.
- Plot the regression line and provide **graphical interpretations** of $\hat{\beta}_1$ and $\hat{\beta}_2$.
- Calculate the residual vector e, mean residual $\bar{e}$ and residual variance $Var(e)$.

OLS: Gauss-Markov Assumptions

The statistical model retains its meaning by assuming some **properties of the error term**. Gauss and Markov did that for us.

$$\begin{aligned}
\text{Linear relationship: } & y_i = x_i'\beta + \epsilon_i \text{ (GM0)} \\
\text{Errors are zero on average: } & E(\epsilon_i) = 0 \forall i \in N \text{ (GM1)} \\
\text{Independence: } & E(\epsilon_i \mid x_i) = E(\epsilon_i) = 0 \forall i \in N \text{ (GM2)} \\
\text{Homoskedasticity: } & Var(\epsilon_i) = \sigma_\epsilon^2 \forall i \in N \text{ (GM3)} \\
\text{No Autocorrelation: } & Cov(\epsilon_i, \epsilon_j) = 0 \forall i \neq j \in N \text{ (GM4)}
\end{aligned}$$

OLS: Unbiasedness

An estimator is unbiased if the expected value of the estimator is the "true" estimator: $E(\hat{\beta}) = \beta$.

$$\begin{aligned} E[\hat{\beta}] = E[(X'X)^{-1}X'y] = \\ E[\beta + (X'X)^{-1}X'\epsilon] = \\ \beta + E[(X'X)^{-1}X'\epsilon] = \beta \end{aligned} \tag{28}$$

Because $E[\epsilon_i \mid x_i) = E[\epsilon_i]$ and $E[\epsilon_i] = 0$, ie. **Gauss-Markov assumptions 1 and 2**.

OLS: Efficiency

An estimator is efficient if it has the **smallest expected variance** within its class of estimators ($Var(\hat{\beta}) \leq Var(\beta^*) \forall \beta^* \neq \hat{\beta}$).

Remember: The variance is the **average squared deviation from the mean** $Var(\beta) = E[(\hat{\beta} - \beta)^2]$.

$$V(\hat{\beta}) = \sigma^2 (X'X)^{-1} = \sigma^2 \left(\sum_{i=1}^{N} (x_i x_i') \right)^{-1} \tag{29}$$

$$\begin{gathered} V(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = E[(X'X)^{-1} X' \epsilon \epsilon' X (X'X)^{-1}] = \\ (X'X)^{-1} X' (\sigma^2 I_N) X (X'X)^{1} = \sigma^2 (X'X)^{-1} \end{gathered} \tag{30}$$

Under Gauss-Markov Conditions 1 through 4 $\hat{\beta}_{OLS} = (X'X)^{-1} X'y$ is the **best unbiased linear estimator (BLUE)**.

OLS: Economics

OLS estimates the marginal impact of a *ceteris paribus* shift in the mean covariate $\bar{x}$ on the mean dependent variable $\bar{y}$.

Linearization, eg. by logarithm, is a popular tool in mainstream economics to apply OLS.

$$
\begin{aligned}
Y &= K^{\alpha} L^{(}\beta) \\
log(Y) &= \alpha log(K) + \beta log(L) \\
log(Y_i) &= \alpha log(K_i) + \beta log(L_i) + \epsilon_i
\end{aligned}
$$

OLS: Goodness of Fit (R^2)

How much of the data variance is explained by the estimated model? R^2 estiamtes the ratio of explained variation in total variation, or 1 minus the sum of squared residuals (SSR) over the sum of squared totals (SST).

$$R^2 = \frac{\hat{V}[\hat{y}]}{\hat{V}[y]} = \frac{1/(N) - 1 \sum(\hat{y} - \bar{y})^2}{1/(N-1)\sum(y_i - \bar{y})^2} \tag{31}$$

If the model **contains an intercept**:

$$y_i = \hat{y}_i + e_i \tag{32}$$

$$\hat{V}(y_i) = \hat{V}(y_i) + \hat{V}(e_i) \tag{33}$$

$$R^2 = 1 - \frac{\hat{V}(e_i)}{\hat{V}(y_i)} = 1 - \frac{1/(N-1)\sum e_i^2}{1/(N-1)\sum(y_i - \bar{y})^2} \tag{34}$$

The **adjusted** R^2 punishes including too many variables by replacing $1/(N-1)$ by $1/(N-K)$.

$$\tilde{R}^2 = 1 - \frac{1/(N-K)\sum e_i^2}{1/(N-K)\sum(y_i - \bar{y})^2} \tag{35}$$

OLS: Significance

Significance: Was it **actually necessary** to include that covariate $\Rightarrow$ Is the corresponding coefficient different from zero?

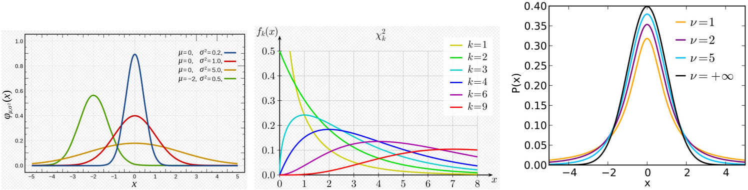
Compare a **test statistic** with known **critical values** (for coefficient estimate $\hat{\beta}$ and H0 value β_k^0). Often, β_k^0 will be zero.

$$t_k = \frac{\hat{\beta}_k - \beta_k^0}{s.e.(\hat{\beta}_k)} \tag{36}$$

If ϵ is **Gaussian Normal distributed** with mean β and variance $\sigma^2(X'X)^{-1}$, and the unknown σ is estimated by s, the unbiased estimator s^2 is χ^2-distributed with $N - K$ degrees of freedom.

Then t_k is the ratio of a Gaussian Normal distribution over the square root of a χ^2 distributed variable, which is $Student - t$ distributed with $N - K$ degrees of freedom.

OLS: Significance Testing Distributions



Wikipedia: Gaussian Normal, χ^2 and Student's t-distributions.

Note that a Student' t distribution with large degrees of freedom (many observations, few covariates) closely resembles a Gaussian Normal.

OLS: Significance Probabilities

Significance testing aks the **probability** that a given estimate $\hat{\beta}$ is the same as some H0 value β^0.

This is equivalent to asking if the corresponding test statistic is higher than some critical value for a given probability α (e.g. for $\alpha = 0.05$).

$$t_{N-K,\alpha/2} s.t. P(\mid t_k \mid > t_{N-K,\alpha/2}) = \alpha \tag{37}$$

For example, for a high $N - K$, the Student's t distribution resembles a Standard Gaussian Normal distribution, and we reject the hypothesis $\hat{\beta}_k = \beta_k^0$ if $t_k > 1.64$.

Example: Gender Wage Gap Summary Statistics

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

Gender	Mean Log Wage
F	6.255308
M	6.729774

Example: Gender Wage Gap Regression

$$
\begin{aligned}
g \in G &= (F, M) \\
w_g &= f(BP, DC) \qquad (38) \\
BP &= f(OS, AC) \qquad (39) \\
w_i &\approx (DC + OS + AC)_i + AC_{i(g)} + \epsilon_i \qquad (40)
\end{aligned}
$$

Example: Gender Wage Gap Regression

```
psid <- foreign::read.dta("../data/mus08psidextract.dta")
options(scipen = 4)
lm(lwage ~ fem, data=psid) %>%
  summary()
```

```
##
## Call:
## lm(formula = lwage ~ fem, data = psid)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.71249 -0.27615  0.01429  0.25402  1.80723
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.72977    0.00718  937.29   <2e-16 ***
## fem         -0.47447    0.02140  -22.18   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4365 on 4163 degrees of freedom
## Multiple R-squared:  0.1056, Adjusted R-squared:  0.1054
## F-statistic: 491.7 on 1 and 4163 DF,  p-value: < 2.2e-16
```

Hotwash

- What did not work today?
- What could I have done better?
- Is there anything you wish you had done differently?

Anonymous Submissions: https://pad.riseup.net/p/ebols-fall2020

Next Lecture: Time Series

Please Read: **Verbeek, 2005, Chapters 8 and 9**