USC, Fall 2016, Economics 513

Lecture 2: Some properties of the OLS solution

Properties of the OLS solution

We fit a linear relation between a dependent variable y and K independent variables $x_1, \ldots, x_K$. The relation also has an intercept. The observations on the dependent variable are in the $n \times 1$ vector y and the observations on the K independent variables are in the $n \times (K+1)$ matrix X with a first column that consist of 1-s.

For the OLS 'estimator'

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

define ($x_{i0} = 1, i = 1, \ldots, n$)

- Vector of OLS residuals

$$e = y - X\hat{\beta} = \begin{bmatrix} y_1 - \sum_{k=0}^{K} x_{1k}\hat{\beta}_k \\ \vdots \\ y_n - \sum_{k=0}^{K} x_{nk}\hat{\beta}_k \end{bmatrix}$$

  This is an $n \times 1$ vector.

- Vector of OLS fitted or predicted values

$$\hat{y} = X\hat{\beta} = \begin{bmatrix} \sum_{k=0}^{K} x_{1k}\hat{\beta}_k \\ \vdots \\ \sum_{k=0}^{K} x_{nk}\hat{\beta}_k \end{bmatrix}$$

  This is an $n \times 1$ vector.

Properties of residuals and fitted values

Property 1:

$$X'e = 0$$

Proof:

$$X'e = X'(y - X\hat{\beta}) = X'y - X'X\hat{\beta} = 0$$

because this is the first order condition for a minimum of $S(\beta)$.

Remarks

- Note if $e = y - X\hat{\beta}$, then $X'e = 0 \Leftrightarrow \hat{\beta} = (X'X)^{-1}X'y$. We could have used $X'e = 0$ to derive the OLS solution $\hat{\beta}$. We come back to this in the CLR model.

- If we partition

$$X = \begin{bmatrix} \iota & x_1 & \cdots & x_K \end{bmatrix}$$

with x_k the $n \times 1$ vector of observations on the k-th explanatory variable and ι an $n \times 1$ vector of 1-s, then (when we transpose the partitioned matrix the first row of the transposed matrix is the first column of the original matrix transposed)

$$X'e = \begin{bmatrix} \iota & x_1 & \cdots & x_K \end{bmatrix}' e = \begin{bmatrix} \iota' \\ x_1' \\ \vdots \\ x_K' \end{bmatrix} e = \begin{pmatrix} \iota' e \\ x_1' e \\ \vdots \\ x_K' e \end{pmatrix}$$

so that $X'e = 0$ implies that

$$\iota' e = \sum_{i=1}^{n} e_i = \bar{e} = 0$$

$$x_k' e = \sum_{i=1}^{n} x_{ik} e_i = 0 \quad , k = 1, \ldots, K$$

In the language of matrix/linear algebra: the vector e is orthogonal to the vectors $\iota, x_1, \ldots, x_K$.

Note

$$e = y - X\hat{\beta} = y - X(X'X)^{-1}X'y = Iy - X(X'X)^{-1}X'y =$$
$$(I - X(X'X)^{-1}X')y = My$$

with

$$M = I - X(X'X)^{-1}X'$$

The $n \times n$ matrix M has a number of properties (see Appendix)

- $MX = 0$
- $M' = M$, i.e. M is symmetric
- $M^2 = M$, i.e. M is idempotent

Also

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

with

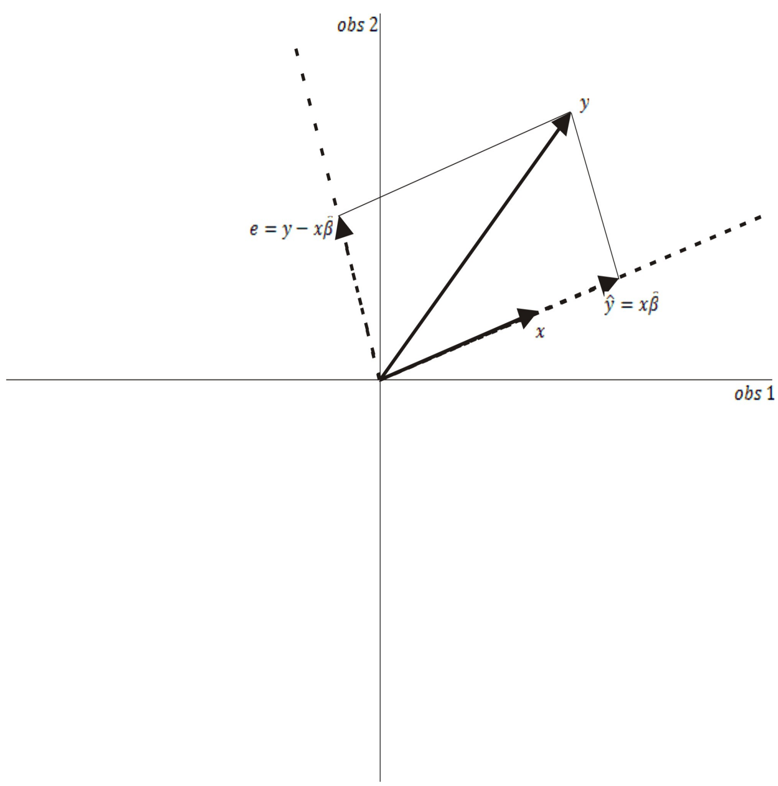
$$P = X(X'X)^{-1}X'$$

Note

- $P' = P$, i.e. P is symmetric
- $P^2 = P$, i.e. P is idempotent

A square symmetric and idempotent matrix is called a projection matrix. The projection $\hat{y}$ of a vector y on a set is the vector in that set that is closest to y. The set is that of all linear combinations of the columns of the matrix X, i.e. the set of $n \times 1$ vectors $v = Xb$ for any $(K+1) \times 1$ vector b. If we take as distance between y and v the Euclidean distance $\sqrt{(y-v)'(y-v)} = \sqrt{(y-Xb)'(y-Xb)}$, then $\hat{y} = X\hat{\beta}$ minimizes this distance (we know that the distance is minimized if $b = \hat{\beta}$) and is the (least squares) projection of y on the set that we call the space spanned by the columns of X.

The projection is illustrated in the figure. We consider a relation between a dependent variable y and one independent variable x. The relation has no intercept, i.e. $y = \beta x + e$ with e the deviation from the line (through the origin). We have two observations on y, denoted by y_1, y_2 and x,denoted by x_1, x_2. The OLS estimator is $\hat{\beta} = \frac{x_1 y_1 + x_2 y_2}{x_1^2 + x_2^2}$.



Obviously the projection of $\hat{y}$ on the same set is $\hat{y}$ and for this reason projection matrices are idempotent. The OLS residual vector e is obtained by projection of y on the space spanned by the vectors that are orthogonal to X (see the figure). The matrix M is the corresponding projection matrix.

Note

$$y = \hat{y} + e$$

with

$$\hat{y}'e = \hat{\beta}'X'e = 0$$

i.e. y can be expresses as the sum of two orthogonal projections, one on the space spanned by the columns of X and one on the space spanned by the vectors that are orthogonal to the columns of X.

Property 2: If the relation has an intercept, then

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

with

$$\bar{x} = \begin{bmatrix} \frac{1}{n}\sum\limits_{i=1}^{n} x_{i0}(=1) \\ \vdots \\ \frac{1}{n}\sum\limits_{i=1}^{n} x_{iK} \end{bmatrix}$$

the $(K+1)\times 1$ vector of sample means of the independent variables. In words: the sample averages of the dependent and independent variables are on the OLS 'line'.

Proof: If we partition X as

$$X = \begin{bmatrix} \iota & x_1 & \cdots & x_K \end{bmatrix}$$

then the normal equations can be written as

$$\begin{bmatrix} \iota' \\ x_1' \\ \vdots \\ x_K' \end{bmatrix} X\hat{\beta} = \begin{bmatrix} \iota' \\ x_1' \\ \vdots \\ x_K' \end{bmatrix} y$$

The first equation is

$$\iota' X\hat{\beta} = \iota' y$$

or

$$\frac{1}{n}\iota' X\hat{\beta} = \frac{1}{n}\iota' y$$

or

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

Here we use

$$\frac{1}{n}\iota' X = \frac{1}{n}\iota' \left[\begin{array}{cccc} \iota & x_1 & \cdots & x_K \end{array} \right] = \left(\begin{array}{cccc} \frac{1}{n}\iota'\iota & \frac{1}{n}\iota' x_1 & \cdots & \frac{1}{n}\iota' x_K \end{array} \right)$$

and e.g.

$$\frac{1}{n}\iota' x_1 = \frac{1}{n}\sum_{i=1}^{n} x_{i1} = \overline{x}_1$$

Therefore

$$\frac{1}{n}\iota' X = \overline{x}'$$

i.e. the $1 \times (K+1)$ row vector of sample means of the columns of X.

Property 3: If the relation has an intercept

$$\bar{y} = \bar{\hat{y}}$$

In words: the sample average of the dependent variable is predicted exactly.

Proof: Because $y = \hat{y} + e$

$$\frac{1}{n}\iota' y = \frac{1}{n}\iota'\hat{y} + \frac{1}{n}\iota' e$$

Property 4: If the relation has an intercept

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} e_i^2$$

Proof: Because $y_i - \bar{y} = \hat{y}_i - \bar{y} + e_i$

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y} + e_i)^2 =$$

$$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} e_i^2 + 2\sum_{i=1}^{n} e_i(\hat{y}_i - \bar{y})$$

The last term is 0.

Terminology

Total Sum of Squares (TSS)

$$\sum_{i=1}^{n}(y_i - \bar{y})^2$$

Explained Sum of Squares (ESS)

$$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

Residual Sum of Squares (RSS)

$$\sum_{i=1}^{n} e_i^2$$

Hence

$$\text{TSS} = \text{ESS} + \text{RSS}$$

If we divide by TSS

$$\frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

Coefficient of determination

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

This is a measure of goodness-of-fit.

To understand this note first

$$0 \leq R^2 \leq 1$$

For the extreme values 0 and 1

$$R^2 = 1 \Leftrightarrow \sum_{i=1}^{n} e_i^2 = 0 \Leftrightarrow e_i = 0, i = 1, \ldots, n \Leftrightarrow y = X\hat{\beta}$$

This means a perfect fit: the observations satisfy the linear relation exactly.

$$\begin{aligned} R^2 = 0 &\Leftrightarrow \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = 0 \Leftrightarrow \hat{y}_i = \bar{y} \Leftrightarrow \\ &\Leftrightarrow \hat{\beta}_0 = \bar{y}, \hat{\beta}_1 = \cdots = \hat{\beta}_K = 0 \end{aligned}$$

i.e. varying regressors $x_1, \ldots x_K$ do not have an (estimated) effect on the dependent variable.

Partitioned regression

Let us consider the linear relation between y and K independent variables $x_1, \ldots x_K$. We split the independent variables in two groups: group 1 has $x_1, \ldots x_{K_1}$ and group two the remaining $K_2 = K - K_1$ variables.

Can we compute the OLS estimates of the coefficients on the variables in group 1 in a linear relation between y and all K independent variables?

First partition X as

$$X = [X_1 X_2]$$

with X_1 an $n \times K_1$ and X_2 and $n \times K_2$ matrix. We also partition β as

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \begin{array}{l} K_1 \\ K_2 \end{array}$$

Define (compare with M defined earlier)

$$M_2 = I - X_2(X_2'X_2)^{-1}X_2'$$

This is a projection matrix.

Define

$$M_2 y = y - X_2\hat{\beta}_2^* = y^*$$

with

$$\hat{\beta}_2^* = (X_2'X_2)^{-1}X_2'y$$

In words: y^* is the vector of OLS residuals of the relation between y and the variables in group 2.

Also define

$$X_1^* = M_2 X_1$$

In words: the k-th column of X_1^* is the vector of OLS residuals of the linear relation between the k-th variable in group 1 and the variables in group 2.

It can be shown (exercise) that

$$\hat{\beta}_1 = (X_1^{*'}X_1^*)^{-1}X_1^{*'}y^*$$

In words: the OLS coefficients on the variables in group 1 in a relation between y and all K variables, $\hat{\beta}_1$, is the OLS estimator if we choose as dependent variable the OLS residuals of y in the regression on the variables in group 2 and as independent variables the OLS residuals of the variables in group 1 in the regression on the variables in group 2.

Note the y^*, X_1^* are 'purged' of the effect of the variables in group 2, because OLS residuals are orthogonal to/uncorrelated with the variables in group 2. This is how least squares implements the ceteris paribus condition, i.e. $\hat{\beta}_1$ is the effect of the variables in group 1 'holding the variables in group 2 constant'.

Special cases

- Columns of X_1 and those of X_2 are orthogonal $X_2'X_1 = 0$, so that $M_2X_1 = X_1$ and

  $$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'y$$

  Conclusion: if we omit the variables in group 2 in the relation then the OLS estimates of the coefficients on the variables in group 1 are unaffected.

- $X_2 = \iota$, so that

$$M_2 = I - \iota(\iota'\iota)^{-1}\iota' = I - \frac{\iota\iota'}{n}$$

$$M_2 y = y - \frac{\iota\iota' y}{n} = y - \iota\bar{y}$$

In words: M_2 takes y in deviation from its sample average. Conclusion: if we take both the dependent and the independent variables in deviation from the sample means we obtain OLS estimators of y on the varying regressors, i.e. all independent variables except the intercept (and the same coefficients as with an intercept in the relation).

Appendix

Properties of the projection matrix $M = I - X(X'X)^{-1}X'$. First

$$MX = (I - X(X'X)^{-1}X')X = X - X(X'X)^{-1}X'X = 0$$

by the property of the inverse matrix and the fact that multiplication of a matrix with the identity matrix leaves the matrix unchanged. Second,

$$M' = (I - X(X'X)^{-1}X')' = I' - (X(X'X)^{-1}X')' = I - X((X'X)^{-1})'X' = I - X(X'X)^{-1}X' = M$$

because the transpose of the inverse is the inverse of the transpose and $X'X$ is symmetric. Third,

$$M^2 = M\left(I - X(X'X)^{-1}X'\right) = M - MX(X'X)^{-1}X' = M$$

because $MX = 0$.