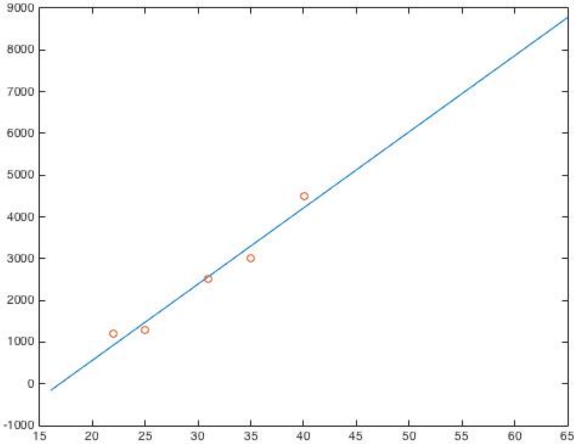
Econometrics - Introduction to Matrix Algebra

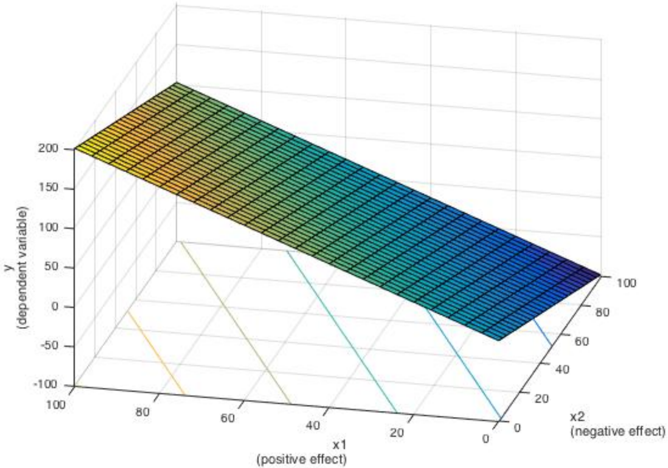
Yale-NUS, YSS2211

February 3, 2016

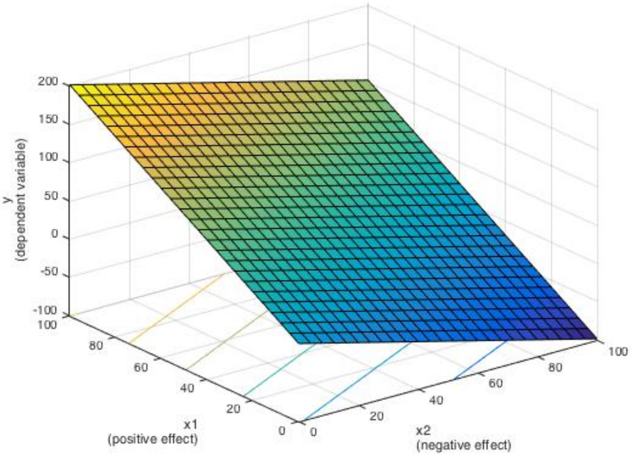
So far: 1 regressor (i.e. $y = f(x)$)



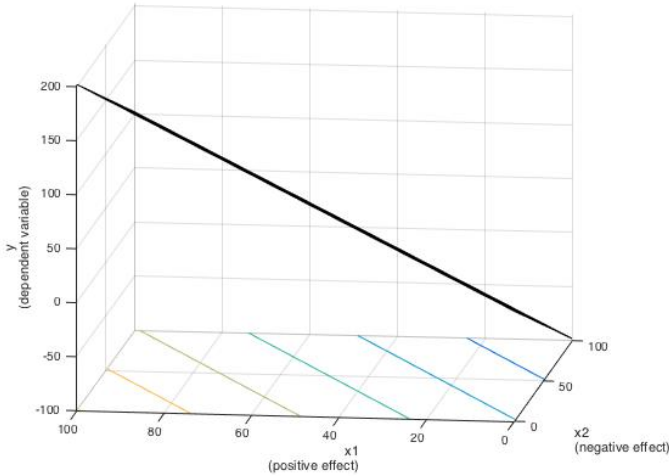
Next: 2 regressors (i.e. $y = f(x_1, x_2)$)



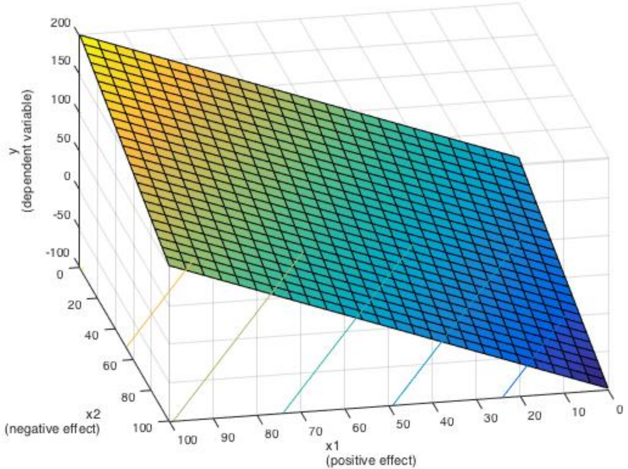
Next: 2 regressors (i.e. $y = f(x_1, x_2)$)



Next: 2 regressors (i.e. $y = f(x_1, x_2)$)



Next: 2 regressors (i.e. $y = f(x_1, x_2)$)



Next: 2 regressors (i.e. $y = f(x_1, x_2)$)

Note: last four slides it's the same graph, just different angles
The model is:
$y = \alpha + \beta_1 x_1 - \beta_2 x_2$
and to be precise, the four previous slides represent:
$y = 2 + 2x_1 - x_2$

So far

$$y = \alpha + \beta x + \varepsilon \tag{1}$$

So far

in more detail

$$\begin{pmatrix} y_1 = \alpha + \beta x_1 + \varepsilon_1 \\ y_2 = \alpha + \beta x_2 + \varepsilon_2 \\ y_3 = \alpha + \beta x_3 + \varepsilon_3 \\ \vdots \\ y_N = \alpha + \beta x_N + \varepsilon_N \end{pmatrix} \tag{2}$$

So far

which in matrix algebra is

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} \alpha \\ \alpha \\ \alpha \\ \vdots \\ \alpha \end{pmatrix} + \begin{pmatrix} x_1\beta \\ x_2\beta \\ x_3\beta \\ \vdots \\ x_N\beta \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_N \end{pmatrix} \tag{3}$$

So far

which in matrix algebra is

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \alpha + \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{pmatrix} \beta + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_N \end{pmatrix} \tag{4}$$

That is,

$$\vec{y} = \alpha + \beta\vec{x} + \vec{\varepsilon} \tag{5}$$

Let's add one more explanatory variable

For instance, wage depends on education (x_1) and age (X_2)

$$\begin{pmatrix} y_1 = \alpha + \beta_1 x_{1,1} + \beta_2 x_{2,1} + \varepsilon_1 \\ y_2 = \alpha + \beta_1 x_{1,2} + \beta_2 x_{2,2} + \varepsilon_2 \\ y_3 = \alpha + \beta_1 x_{1,3} + \beta_2 x_{2,3} + \varepsilon_3 \\ \vdots \\ y_N - \alpha + \beta_1 x_{1,N} + \beta_2 x_{2,N} + \varepsilon_N \end{pmatrix} \tag{6}$$

Let's add one more explanatory variable

We want to estimate α, β_1, β_2. Like in the case with only one regressor, we will minimize the sum of squared errors.

$$\begin{pmatrix} y_1 - \alpha + \beta_1 x_{1,1} + \beta_2 x_{2,1} = \varepsilon_1 \\ y_2 - \alpha + \beta_1 x_{1,2} + \beta_2 x_{2,2} = \varepsilon_2 \\ y_3 - \alpha + \beta_1 x_{1,3} + \beta_2 x_{2,3} = \varepsilon_3 \\ \vdots \\ y_N - \alpha + \beta_1 x_{1,N} + \beta_2 x_{2,N} = \varepsilon_N \end{pmatrix} \tag{7}$$

Our problem

We want to estimate α, β_1, β_2. Like in the case with only one regressor, we will minimize the sum of squared errors.

$$Min_{\alpha,\beta_1,\beta_2} \sum_{i=1}^{N} \varepsilon_i^2 = Min_{\alpha,\beta_1,\beta_2} \sum_{i=1}^{N} (y_i - \alpha + \beta_1 x_{1,i} + \beta_2 x_{2,i})^2 \quad (8)$$

First Order Conditions

$$
\begin{aligned}
[\alpha] \qquad & N\hat{\alpha} + \hat{\beta}_1 \sum_{i=1}^{N} x_{1,i} + \hat{\beta}_2 \sum_{i=1}^{N} x_{2,i} = \sum_{i=1}^{N} y_i \\
[\beta_1] \qquad & \hat{\alpha} \sum_{i=1}^{N} x_{1,i} + \hat{\beta}_1 \sum_{i=1}^{N} x_{1,i}^2 + \hat{\beta}_2 \sum_{i=1}^{N} x_{1,i} x_{2,i} = \sum_{i=1}^{N} x_{1,i} y_i \qquad (9) \\
[\beta_2] \qquad & \hat{\alpha} \sum_{i=1}^{N} x_{2,i} + \hat{\beta}_1 \sum_{i=1}^{N} x_{1,i} x_{2,i} + \hat{\beta}_2 \sum_{i=1}^{N} x_{2,i}^2 = \sum_{i=1}^{N} x_{2,i} y_i
\end{aligned}
$$

First Order Conditions

This is t-e-d-i-o-u-s!

Solution: Matrix Algebra

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} \alpha \\ \alpha \\ \alpha \\ \vdots \\ \alpha \end{pmatrix} + \begin{pmatrix} x_{1,1}\beta_1 \\ x_{1,2}\beta_1 \\ x_{1,3}\beta_1 \\ \vdots \\ x_{1,N}\beta_1 \end{pmatrix} + \begin{pmatrix} x_{2,1}\beta_2 \\ x_{2,2}\beta_2 \\ x_{2,3}\beta_2 \\ \vdots \\ x_{2,N}\beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_N \end{pmatrix} \tag{10}$$

Solution: Matrix Algebra

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \alpha + \begin{pmatrix} x_{1,1} \\ x_{1,2} \\ x_{1,3} \\ \vdots \\ x_{1,N} \end{pmatrix} \beta_1 + \begin{pmatrix} x_{2,1} \\ x_{2,2} \\ x_{2,3} \\ \vdots \\ x_{2,N} \end{pmatrix} \beta_2 + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_N \end{pmatrix} \tag{11}$$

Solution: Matrix Algebra

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} 1 & x_{1,1} & x_{2,1} \\ 1 & x_{1,2} & x_{2,1} \\ 1 & x_{1,3} & x_{2,1} \\ \vdots & & \\ 1 & x_{1,N} & x_{2,1} \end{pmatrix}}_{X} \underbrace{\begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{pmatrix}}_{\beta} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_N \end{pmatrix}}_{\varepsilon} \tag{12}$$

Solution: Matrix Algebra

$$y = X\beta + \varepsilon \tag{13}$$

Note, using Matrix properties,

$$y_1 = 1 \times \alpha + x_{1,1} \times \beta_1 + x_{2,1} \times \beta_2 + \varepsilon_1$$

and generally, for any observation i in our sample,

$$y_i = \alpha + x_{1,i} \times \beta_1 + x_{2,i} \times \beta_2 + \varepsilon_i$$

Our Problem is the same as above (see equation (7))

We want to estimate α, β_1, β_2. Like in the case with only one regressor, we will minimize the sum of squared errors.

$$\begin{pmatrix} y_1 - \alpha + \beta_1 x_{1,1} + \beta_2 x_{2,1} = \varepsilon_1 \\ y_2 - \alpha + \beta_1 x_{1,2} + \beta_2 x_{2,2} = \varepsilon_2 \\ y_3 - \alpha + \beta_1 x_{1,3} + \beta_2 x_{2,3} = \varepsilon_3 \\ \vdots \\ y_N - \alpha + \beta_1 x_{1,N} + \beta_2 x_{2,N} = \varepsilon_N \end{pmatrix}$$

but now it looks nicer:
$y - X\beta = \varepsilon$. Our problem is to find the argument that minimizes

$$Min \quad \varepsilon'\varepsilon = Min\,(y - X\beta)'\,(y - X\beta)$$

Vector multiplication

Why do we write

$$Min \quad \varepsilon'\varepsilon = Min\,(y - X\beta)'\,(y - X\beta)$$

instead of

$$Min \quad \varepsilon^2 = Min\,(y - X\beta)^2 \quad ?$$

This is a useful convention. We cannot square or cube matrices, but we can always pre-multiply them by themselves, or post-multiply them by themselves

Vector multiplication: What is $\varepsilon'\varepsilon$?

$$\varepsilon'\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_N \end{pmatrix}^T \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

i.e.

$$\varepsilon'\varepsilon = \underbrace{\begin{pmatrix} \varepsilon_1 & \varepsilon_2 & \varepsilon_3 & \dots & \varepsilon_N \end{pmatrix}}_{1 \times N} \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_N \end{pmatrix}}_{N \times 1} = \varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2 + \dots + \varepsilon_n^2 = \sum_{i=1}^{N} \varepsilon_i^2$$

Note: $\varepsilon'\varepsilon \neq \varepsilon\varepsilon'$!

$$\varepsilon\varepsilon' = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_N \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_N \end{pmatrix}^T$$

i.e.

$$\varepsilon\varepsilon' = \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_N \end{pmatrix}}_{N \times 1} \underbrace{\begin{pmatrix} \varepsilon_1 & \varepsilon_2 & \varepsilon_3 & \dots & \varepsilon_N \end{pmatrix}}_{1 \times N} = \begin{pmatrix} \varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \varepsilon_1\varepsilon_3 & \dots & \varepsilon_1\varepsilon_N \\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & \varepsilon_2\varepsilon_3 & \dots & \varepsilon_2\varepsilon_N \\ \varepsilon_3\varepsilon_1 & \varepsilon_3\varepsilon_2 & \varepsilon_3^2 & \dots & \varepsilon_3\varepsilon_N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \varepsilon_N\varepsilon_1 & \varepsilon_N\varepsilon_2 & \varepsilon_N\varepsilon_3 & \dots & \varepsilon_N^2 \end{pmatrix}$$

Our Problem is the same as above (see equation (7))

We want to estimate α, β_1, β_2 in $y = X\beta + \varepsilon$. Like in the case with only one regressor, we will minimize the sum of squared errors. That is, our problem is to find the argument that minimizes

$$Min_{\beta} \quad \varepsilon'\varepsilon = Min\,(y - X\beta)'\,(y - X\beta)$$

where β is a vector of parameters: $\beta = \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{pmatrix}$

OLS in Matrix Algebra

$$
\begin{aligned}
& Min_\beta \quad \varepsilon'\varepsilon = Min\,(y - X\beta)'\,(y - X\beta) \\
& = Min\left(y' - \beta'X'\right)(y - X\beta) \\
& = Min\left(y'y - \beta'X'y - y'X\beta + \beta'X'X\beta\right) \\
& = Min\left(y'y - 2\beta'X'y + \beta'X'X\beta\right)
\end{aligned}
$$

where the last step is due to the fact that $\beta'X'y$ is a scalar: $(1 \times k)(k \times N)(N \times 1)$, and we know $\lambda' = \lambda$ for any scalar λ
Also note: $y'y$ is a scalar, and $\beta'X'X\beta$ is also a scalar

Matrix Differentiation

Let c be a $k \times 1$ vector: $c = \begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix}$

Let β be a $k \times 1$ vector of parameters: $\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$

$c'\beta = ?$

Matrix Differentiation

Let c be a $k \times 1$ vector: $c = \begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix}$

Let β be a $k \times 1$ vector of parameters: $\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$

$c'\beta = ?\ \ c_1\beta_1 + c_2\beta_2 + \ldots + c_k\beta_k$

Matrix Differentiation

What is then $\frac{\partial(c'\beta)}{\partial\beta}$?

$$\frac{\partial(c'\beta)}{\partial\beta} = \begin{pmatrix} \frac{\partial(c_1\beta_1+c_2\beta_2+\ldots+c_k\beta_k)}{\partial\beta_1} \\ \frac{\partial(c_1\beta_1+c_2\beta_2+\ldots+c_k\beta_k)}{\partial\beta_2} \\ \vdots \\ \frac{\partial(c_1\beta_1+c_2\beta_2+\ldots+c_k\beta_k)}{\partial\beta_k} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} = c$$

Note: since $c'\beta = \beta'c \Rightarrow \frac{\partial(\beta'c)}{\partial\beta} = c$

OLS in Matrix Algebra: FOC [$\frac{\partial(\beta' c)}{\partial\beta} = c$ and $\frac{\partial(c'\beta)}{\partial\beta} = c$]

$$\frac{\partial\left(y'y - 2\beta' X' y + \beta' X' X \beta\right)}{\partial\beta}$$

$$= -2X'y + X'X\beta + (\beta' X' X)' =$$

$$= -2X'y + X'X\beta + (X'X\beta) =$$

$$= -2X'y + 2X'X\beta = 0$$

$$\Rightarrow X'y = X'X\beta$$

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

OLS in Matrix Algebra: FOC

Note: $(X'X)^{-1}X'y = \hat{\beta}$ looks very much like $\frac{\sigma_{xy}}{\sigma_x^2} = \hat{\beta}$