Mathematics for Economists
Aalto BIZ
Spring2021
Juuso Välimäki

Economic Applications

Ordinary least squares

Consider a statistical sample consisting of on N pairs of observations

$$(y_1, x_1), ..., (y_N, x_N).$$

Suppose that we want to find a linear relation between x and y. We would like to find a coefficient β that rationalizes the observations as

$$y_i = \beta x_i.$$

If we have many observations, this will not be satisfied in general. To account for errors in the linear relationship, specify the following statistical model

$$y_i = \beta x_i + \varepsilon_i,$$

where ε_i is an identically and independently distributed error term for all i.

Our task is to infer β from the sample. One way of doing this is based on minimizing the sum of squared error terms $(\varepsilon_i)^2$, i.e. to

$$\min_{\beta} f(\beta) = \Sigma_{i=1}^{N} (y_i - \beta x_i)^2.$$

Compute $f'(\beta)$ and consider $\widehat{\beta}$ such that:

$$f'\left(\widehat{\beta}\right) = 0.$$

By taking the derivative, we get:

$$f'\left(\widehat{\beta}\right) = \sum_{i=1}^{N} -2x_i \left(y_i - \widehat{\beta} x_i\right).$$

As a result, $f'\left(\widehat{\beta}\right) = 0$ if

$$\widehat{\beta} = \frac{\Sigma_{i=1}^{N} x_i y_i}{\Sigma_{i=1}^{N} x_i^2}.$$

Since $f''(\hat{\beta}) = \sum_{i=1}^{N} x_i^2 > 0$, we have found the minimum.

If we want to include a constant term α, we get:

$$y_i = \alpha + \beta x_i + \varepsilon_i.$$

function of (α, β):

$$f(\alpha, \beta) = \Sigma_{i=1}^{N} (y_i - \alpha - \beta x_i)^2.$$

To find $\left(\widehat{\alpha}, \widehat{\beta}\right)$ such that

$$\frac{\partial f\left(\widehat{\alpha}, \widehat{\beta}\right)}{\partial \alpha} = \frac{\partial f\left(\widehat{\alpha}, \widehat{\beta}\right)}{\partial \beta} = 0,$$

we get:

$$\begin{aligned} \Sigma_{i=1}^{N} \left(y_i - \widehat{\alpha} - \widehat{\beta} x_i\right) &= 0, \\ \Sigma_{i=1}^{N} - 2x_i \left(y_i - \widehat{\alpha} - \widehat{\beta} x_i\right) &= 0. \end{aligned}$$

Solving for α from the first equation gives:

$$\widehat{\alpha} = \frac{\Sigma_{i=1}^{N} y_i - \widehat{\beta} \Sigma_{i=1}^{N} x_i}{N} := \overline{y} - \widehat{\beta} \overline{x}.$$

Using the first equation we also see that:

$$\Sigma_{i=1}^{N} \overline{x} \left(y_i - \widehat{\alpha} - \widehat{\beta} x_i\right) = 0.$$

By substituting into the second, we get:

$$\widehat{\beta} = \frac{\Sigma_{i=1}^{N} (x_i - \overline{x})(y_i - \overline{y})}{\Sigma_{i=1}^{N} (x_i - \overline{x})^2} = \frac{Cov(y, x)}{Var(x)}.$$

More generally, we can consider samples with more explanatory variables: $(y_1, x_{11}, x_{21}, ..., x_{K1}), ... (y_N, x_{1N}, ..., x_{KN})$ and a linear model

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \begin{pmatrix} = \\ \vdots \\ = \end{pmatrix} \begin{pmatrix} \beta_1 x_{11} + & \cdots & \beta_K x_{K1} \\ \vdots & & \vdots \\ \beta_1 x_{1N} & \cdots & \beta_K x_{KN} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

or in matrix form:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}.$$

We can compute the sum of squares now as:

$$\begin{aligned} f(\boldsymbol{\beta}) &= \boldsymbol{\varepsilon} \cdot \boldsymbol{\varepsilon} = (\boldsymbol{y} - \boldsymbol{X\beta})^\top (\boldsymbol{y} - \boldsymbol{X\beta}) \\ &= \boldsymbol{y} \cdot \boldsymbol{y} - (\boldsymbol{X\beta})^\top \boldsymbol{y} - \boldsymbol{y}^\top \boldsymbol{X\beta} + \boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X\beta} \\ &= \boldsymbol{y} \cdot \boldsymbol{y} - 2\boldsymbol{y}^\top \boldsymbol{X\beta} + \boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X\beta}. \end{aligned}$$

The derivative of $-2\boldsymbol{y}^\top \boldsymbol{X\beta}$ is the row vector $-2\boldsymbol{y}^\top \boldsymbol{X}$ and the derivative of $\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X\beta}$ is $2\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X}$. (To see this, write the matrix product as a sum). Writing with the gradient, we have:

$$\nabla f\left(\widehat{\boldsymbol{\beta}}\right) = -2\boldsymbol{X}^\top \boldsymbol{y} + 2\boldsymbol{X}^\top \boldsymbol{X}\widehat{\boldsymbol{\beta}}.$$

Therefore we can find a candidate for the extremum by setting

$$\nabla f\left(\widehat{\boldsymbol{\beta}}\right) = 0.$$

Solving for $\boldsymbol{\beta}$, we get:

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y}.$$

Exercise: Is there a need to add a constant term to get a more general formula? Show also that the Hessian of $f(\boldsymbol{\beta})$ is positive definite.

Second derivatives of functions used in economics

Power function

Sums of power functions are common in consumer theory. Each component in the sum measures the utility from consumption in a given period.

(Individual components x_i^ρ are often called CRRA functions since they display constant relative risk aversion).

$$f(x_1, x_2) = x_1^\rho + x_2^\rho.$$

Form the gradient

$$\nabla f(x_1, x_2) = \begin{pmatrix} \frac{\partial f(x_1,x_2)}{\partial x_1} \\ \frac{\partial f(x_1,x_2)}{\partial x_2} \end{pmatrix} = \begin{pmatrix} \rho x_1^{\rho-1} \\ \rho x_2^{\rho-1} \end{pmatrix}.$$

Form the Hessian matrix by taking the derivative of the gradient:

$$Hf(x_1, x_2) = \begin{pmatrix} \frac{\partial^2 f(x_1,x_2)}{\partial x_1 \partial x_1} & \frac{\partial^2 f(x_1,x_2)}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f(x_1,x_2)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x_1,x_2)}{\partial x_2 \partial x_2} \end{pmatrix}.$$

We get:

$$Hf(x_1, x_2) = \begin{pmatrix} \rho(\rho-1) x_1^{\rho-2} & 0 \\ 0 & \rho(\rho-1) x_2^{\rho-2} \end{pmatrix}.$$

$D^2 f(x_1, x_2)$ is thus negative definite when $x_i \neq 0$ ja $0 < \rho < 1$.

CES -function

Recall the CES -function (utility and production function).

$$f(x_1, x_2) = (x_1^\rho + x_2^\rho)^{\frac{1}{\rho}}.$$

Form the gradient:

$$\nabla f(x_1, x_2) = \begin{pmatrix} \frac{\partial f(x_1,x_2)}{\partial x_1} \\ \frac{\partial f(x_1,x_2)}{\partial x_2} \end{pmatrix} = \begin{pmatrix} (x_1^\rho + x_2^\rho)^{\frac{1}{\rho}-1} x_1^{\rho-1} \\ (x_1^\rho + x_2^\rho)^{\frac{1}{\rho}-1} x_2^{\rho-1} \end{pmatrix}.$$

The the Hessian matrix is:

$$Hf(x_1, x_2) = \begin{pmatrix} \frac{\partial^2 f(x_1,x_2)}{\partial x_1 \partial x_1} & \frac{\partial^2 f(x_1,x_2)}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f(x_1,x_2)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x_1,x_2)}{\partial x_2 \partial x_2} \end{pmatrix}.$$

By the product rule:

$$\begin{aligned} \frac{\partial^2 f(x_1, x_2)}{\partial x_1 \partial x_1} &= (\rho - 1) x_1^{\rho-2} (x_1^\rho + x_2^\rho)^{\frac{1}{\rho}-1} \\ &\quad + \left(\frac{1}{\rho} - 1\right) (x_1^\rho + x_2^\rho)^{\frac{1}{\rho}-2} \rho x_1^{2\rho-2}, \end{aligned}$$

$$\frac{\partial^2 f(x_1, x_2)}{\partial x_1 \partial x_2} = \left(\frac{1}{\rho} - 1\right) \left(x_1^\rho + x_2^\rho\right)^{\frac{1}{\rho}-2} \rho x_2^{\rho-1} x_1^{\rho-1},$$

$$\begin{aligned} \frac{\partial^2 f(x_1, x_2)}{\partial x_2 \partial x_2} &= (\rho - 1) x_2^{\rho-2} \left(x_1^\rho + x_2^\rho\right)^{\frac{1}{\rho}-1} \\ &\quad + \left(\frac{1}{\rho} - 1\right) \left(x_1^\rho + x_2^\rho\right)^{\frac{1}{\rho}-2} \rho x_2^{2\rho-2}. \end{aligned}$$

By collecting the common terms, we get:

$$D^2 f(x_1, x_2) = \begin{pmatrix} \frac{\partial^2 f(x_1,x_2)}{\partial x_1 \partial x_1} & \frac{\partial^2 f(x_1,x_2)}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f(x_1,x_2)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x_1,x_2)}{\partial x_2 \partial x_2} \end{pmatrix}$$
$$= \left(x_1^\rho + x_2^\rho\right)^{\frac{1}{\rho}-2} \begin{pmatrix} (\rho - 1) x_1^{\rho-2} x_2^\rho & (1 - \rho) x_2^{\rho-1} x_1^{\rho-1} \\ (1 - \rho) x_2^{\rho-1} x_1^{\rho-1} & (\rho - 1) x_2^{\rho-2} x_1^\rho \end{pmatrix}.$$

When computing the determinant, we can separate the common factor:

$$\det\left(D^2 f(x_1, x_2)\right) =$$

$$\left(x_1^\rho + x_2^\rho\right)^{\frac{1}{\rho}-2} x_1^{2\rho-2} x_2^{2\rho-2} \det \begin{pmatrix} (\rho - 1) & (1 - \rho) \\ (1 - \rho) & (\rho - 1) \end{pmatrix} = 0.$$

$D^2 f(x_1, x_2)$ is therefore negative semidefinite if $\rho < 1$ and positive semidefinite if$\rho > 1$.

Definiteness and comparative statics

Consider the unconstrained optimization of choosing $y \in \mathbb{R}$ to reach the highest possible value of:

$$f(y; x),$$

where $x \in \mathbb{R}$ is an exogenous variable. Write the problem of maximizing $y : n$ as follows:

$$\max_y f(y; x)$$

The first order condition for optimum at $(\hat{y}, \hat{x})$ is:

$$\frac{\partial f}{\partial y} (\widehat{y}; \widehat{x}) = 0.$$

A sufficient condition for local maximum is obtained from Taylor's theorem:

$$f(\widehat{y}+dy;\widehat{x})-f(\widehat{y};\widehat{x})=\frac{\partial f}{\partial y}(\widehat{y};\widehat{x})\,dy+\frac{1}{2}\frac{\partial^2 f}{\partial y\partial y}(\widehat{y};\widehat{x})\,(dy)^2+\text{ h.o.t.}$$

If

$$\frac{\partial^2 f}{\partial y\partial y}(\widehat{y};\widehat{x})<0,$$

then $f$has a local maximum at $(\widehat{y};\widehat{x})$.

Note that then also the function

$$\frac{\partial f}{\partial y}(\widehat{y};\widehat{x})$$

has a non-zero derivative w.r.t. the endogenous variable at $(\widehat{y};\widehat{x})$ and we can apply the implicit function theorem y to get the optimal y as a function of x.

Since

$$\frac{\partial f}{\partial y}(y(x);x)=0.$$

for all x near $\widehat{x}$, we get:

$$\frac{\partial^2 f(\widehat{y};\widehat{x})}{\partial y\partial y}dy+\frac{\partial^2 f(\widehat{y};\widehat{x})}{\partial y\partial x}dx=0,$$

or

$$\frac{dy}{dx}=-\frac{\frac{\partial^2 f(\widehat{y};\widehat{x})}{\partial y\partial x}}{\frac{\partial^2 f(\widehat{y};\widehat{x})}{\partial y\partial y}}.$$

Since $\frac{\partial^2 f(\widehat{y};\widehat{x})}{\partial y\partial y}<0$ by second-order condition for optimum, we see that $\frac{dy}{dx}$ has the same sign as $\frac{\partial^2 f(\widehat{y};\widehat{x})}{\partial y\partial x}$.

Example 1 (Optimal monopoly production)**.** Let q be the output by the monopolist. Denote the inverse demand by $p(q;\alpha)$ and assume that it is twice differentiable and downward sloping $\frac{\partial p(q;\alpha)}{\partial q}<0$ for all $q>0$. Let α be a demand shifting variable with such as the income of the consumers and $\frac{\partial p(q;\alpha)}{\partial \alpha}>0$ for all $q>0$.

Denote the cost function by $c(q;\beta)$. Assume that the cost is increasing q and also that the marginal cost is increasing in q. Let β be a cost shifting exogenous variable e.g. price of oil and assume that $\frac{\partial c(q;\beta)}{\partial \beta} < 0$ for all $q > 0$ The monopolist's maximization problem is then:

$$\max_q \pi\left(q;\alpha,\beta\right) = qp(q;\alpha) - c(q;\beta)$$

First-order condition for optimality:

$$D\pi\left(q;\alpha,c\right) = p(q;\alpha) + q\frac{\partial p(q;\alpha)}{\partial q} - \frac{\partial c(q;\beta)}{\partial q} = 0.$$

Second-order condition:

$$H\pi\left(q\right) < 0.$$

If $p'(q;\alpha)$ is decreasing in q, then the second derivative is negative for all q.

How does the optimal output change when α or c changes? By the previous result, the sign of the change in the endogenous variable depends on the signs of

$$\frac{\partial^2 \pi\left(q;\alpha,\right)}{\partial q \partial \alpha}, \frac{\partial^2 \pi\left(q;\alpha,\beta\right)}{\partial q \partial \beta}.$$

Notice that we also need to know the signs of $\frac{\partial^2 p(q;\alpha)}{\partial q \partial \alpha}$ and $\frac{\partial^2 c(q;\beta)}{\partial q \partial \beta}$ to determine how optimal output changes.