FE2713 **Ordinary Least Squares (OLS)**

Ordinary least squares (OLS), also known as the method of least squares, is a technique used to determine how one quantity Y (the dependent variable) changes as another quantity X (the independent variable) is varied, when it is assumed that a linear dependency would adequately describe this relationship. Usually one is in possession of data consisting of pairs of observations (x, y), and the task is to determine the (assumed) linear relationship that holds between X and Y. OLS provides such a linear relationship that is optimal in a sense to be described shortly. The main use of determining this linear relationship is to predict the value of Y on some other X which is not in the data set. An example will illustrate the procedure.

Suppose that hypothetical data relating a persons average investment in mutual funds to the previous year's gross income is available as in Figure 1.

Previous year's income ($\$ \times 1000$)	Investment ($\$ \times 1000$)
25.0	3.2
30.0	2.7
35.0	3.9
40.0	6.9
45.0	5.1
50.0	9.6
55.0	10.1

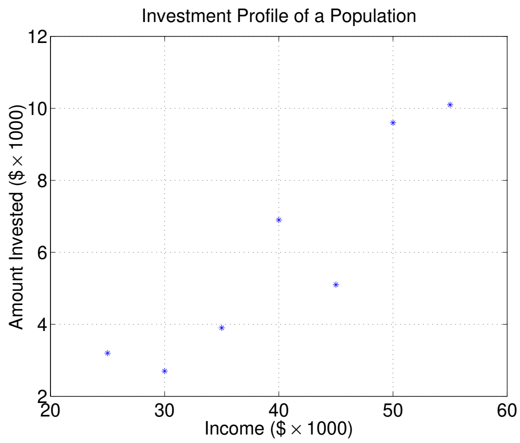

Figure 1. Data describing how investment depends on the previous year's income.

It may be desirable to describe this relationship using a linear dependency,

$$Y = \alpha X + \beta, \tag{1}$$

where X is the previous year's income, and Y is the current year's investment. Intuitively, this corresponds to the expectation that the increase in investment should be proportional to the increase in one's income. Such a linear dependency could, for example, be used to predict the expected demand for mutual funds next year given the estimated increase in incomes for the current year. A hypothetical such linear dependency, often called a linear "fit" to the data, is illustrated in Figure 2(a) (for particular values of α, β), along with the *residuals* illustrated by the vertical bars, which represent errors between the hypothetical linear dependence and the observed data (stars in the figure). There are many such linear fits, and OLS provides a principled way of picking one such fit. The *RMS error* is defined as the square root of the average of the squared errors and the RMS error is indicated in the figures. The OLS linear fit picks

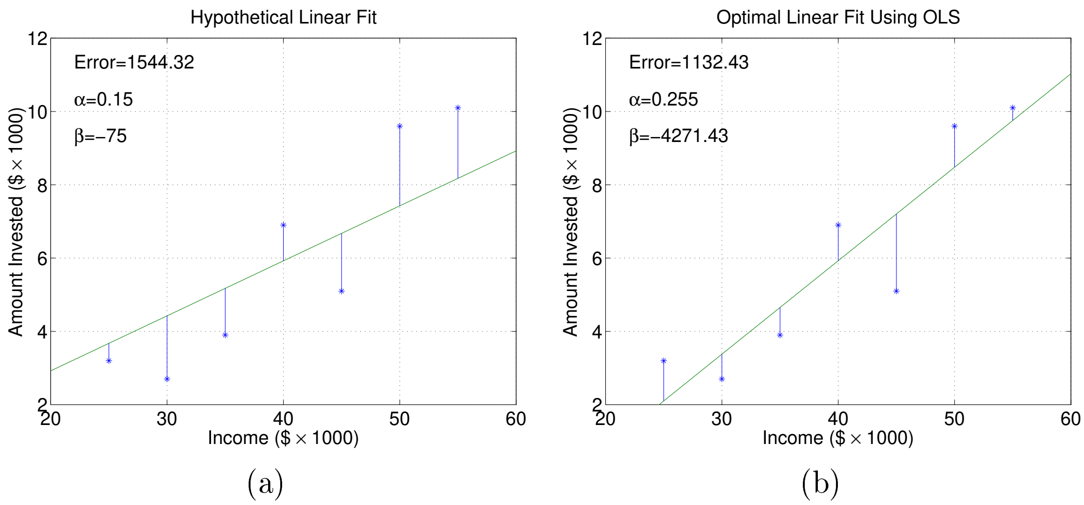

Figure 2. (a) A hypothetical linear fit to the investment data. (b) The OLS linear fit to the investment data. The vertical bars indicate the errors that the linear fit makes, and the error indicated is the RMS error. α and β are the slope and intercept of the fits

the linear dependency (by specifying parameters α, β) that minimizes this RMS error, and is shown in Figure 2(b), along with the RMS error of the OLS linear fit and the parameters, α, β.

The general procedure for determining the OLS linear fit is as follows. Suppose that the data is represented as a set $\{(x_i, y_i)\}$ for $i = 1, 2 \ldots, N$. In the example above, $N = 7$ and for example $(x_2, y_2) = (30000, 2700)$. We denote by $\overline{X}$ and $\overline{Y}$ the average value of the x's and y's respectively, by $\overline{X^2}$ the average value of the x^2's and by $\overline{XY}$ the average value of $x \times y$. Then the OLS parameters α and β are given by (see for example DeGroot and Schervish, 2002, ch. 10)

$$\begin{aligned} \alpha &= \frac{\overline{XY} - \overline{X}\,\overline{Y}}{\overline{X^2} - \overline{X}^2} & \qquad \beta &= \overline{Y} - \alpha\overline{X} \\ &= \frac{cov(X,Y)}{var(X)} \end{aligned} \tag{2}$$

Continuing with the example, we compute the quantities needed to determine α and β.

$$\begin{aligned} \overline{X} &= \frac{10^3 \times (25 + 30 + 35 + 40 + 45 + 50 + 55)}{7} = 40000 \\ \overline{Y} &= \frac{10^3 \times (3.2 + 2.7 + 3.9 + 6.9 + 5.1 + 9.6 + 10.1)}{7} = 5928.57 \\ \overline{XY} &= \frac{10^7 \times (8 + 8.1 + 13.65 + 27.6 + 22.95 + 48 + 55.55)}{7} \\ &= 2.6264 \times 10^8 \\ \overline{X^2} &= \frac{10^8 \times (6.25 + 9 + 12.25 + 16 + 20.25 + 25 + 30.25)}{7} \\ &= 1.7 \times 10^9 \end{aligned} \tag{3}$$

So we find that

$$\begin{aligned} \alpha &= \frac{2.6264 \times 10^8 - 40000 \times 5928.57}{1.7 \times 10^9 - 40000^2} \\ &= 0.255 \\ \\ \beta &= 5928.57 - 0.255 \times 40000 \\ &= -4271.43 \end{aligned} \tag{4}$$

Least squares techniques can be linked to statistical inference through the maximum likelihood framework, when Gaussian noise is assumed (DeGroot and Schervish, 2002) For further details and derivations of the necessary formula, the reader is referred to any one of the books in the further reading. One can generalize the concept of OLS to fitting by polynomials, and to the situation when there are more than one independent variable. These cases are dealt with in the further reading as well as in the section on multiple regression.

MALIK MAGDON-ISMAIL

Further Reading

Anderson, D. R., Sweeney, D. J., and Williams, T. A. (2001). *Statistics for Business and Economics.* South Western Publishers.

DeGroot, M. H. and Schervish, M. J. (2002). *Probability and Statistics.* Addison–Wesley, Reading, Massachusetts.

Devore, J. L. (1999). *Probability and Statistics for Engineering and the Sciences.* Duxbury Press.

Hair, J. F. e., Anderson, R. E., Tatham, R. L., and Black, W. (1998). *Multivariate Data Analysis, 5th Edition.* Prentice Hall, New Jersey.

Larsen, R. and Marx, M. (1986). *An Introduction to Mathematical Statistics.* Prentice-Hall, New Jersey.

Ross, S. M. (2001). *A First Course in Probability.* Prentice Hall, New Jersey.

See also: Multiple regression, Averages (computations).