Lecture 3

Estimation using OLS

Ananya Iyengar

January 2026

St. Stephen's College

This corresponds to Chapter 2 of Studenmund (2017).

The previous lecture established the basic idea of a regression, and provided a distinction between the true relationship between variables in the population and estimating that relationship in the sample (i.e. finding $\hat{Y}$ as an estimate of $E[Y/X]$.

The question remains: how does one estimate a regression?

1 Ordinary Least Squares

1. We want to fit a line through the observed data pairs (X_i, Y_i) – and we would like this fit to be *good*. We noted in the previous lecture that the regression had 2 components: the deterministic and the stochastic random error. Then, a good fit implies wanting to minimise this error.

2. The random error e_i can be positive or negative. Therefore, the Ordinary Least Squares method minimises the sum of the squared values of the error. Formally,

$$\min_\beta \sum_{i=1}^{n} e_i^2 \implies \min_\beta \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2$$

   The first order conditions are given by:

   - Differentiation w.r.t β_0: $2\sum_i (Y_i - \beta_0 - \beta_1 X_i)(-1) = 0$
     $\implies \sum Y_i - n\beta_0 - \beta_1 \sum X_i = 0$

$\implies \bar{Y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{X} = 0$

- Differentiating w.r.t β_1: $2 \sum_i (Y_i - \beta_0 - \beta_1 X_i)(-X_i) = 0$

  $\implies \sum_i (-X_i Y_i + X_i(\bar{Y} - \beta_i \bar{X}) + \beta_1 X_i^2) = 0$

  $\implies -\sum X_i Y_i + (\bar{Y} - \beta_i \bar{X}) \sum X_i + \beta_1 \sum X_i^2 = 0$

  $\implies \beta_1 (\sum X_i^2 - \bar{X} \sum X_i) = \sum X_i Y_i - \bar{Y} \sum X_i$

  $\implies \hat{\beta}_1 = \frac{\sum X_i Y_i - \bar{Y} \sum X_i}{(\sum X_i^2 - \bar{X} \sum X_i)}$

  We can simplify the numerator as follows:

  $\sum X_i Y_i - \bar{Y} \sum X_i = \sum X_i Y_i - n\bar{Y}\bar{X} = \sum X_i Y_i - n\bar{Y}\bar{X} - n\bar{Y}\bar{X} + n\bar{Y}\bar{X}$

  $= \sum (X_i Y_i - \bar{Y} X_i - \bar{X} Y_i + \bar{Y}\bar{X})$

  $= \sum (X_i - \bar{X})(Y_i - \bar{Y})$.

  Similarly, the denominator can be simplified as $\sum (X_i - \bar{X})^2$ (*try this yourself!*)

  $\implies \hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$

3. The sum of OLS residuals $\sum e_i = 0$.

   Proof. $\sum e_i = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$

   $\implies \sum (Y_i - \bar{Y} + \hat{\beta}_1 \bar{X} - \hat{\beta}_1 X_i) = 0$

   $\implies n\bar{Y} - n\bar{Y} + n\hat{\beta}_1 \bar{X} - n\hat{\beta}_1 \bar{X} = 0$ $\square$

4. This same algorithm can be extended to multivariate regression models – all results go through.

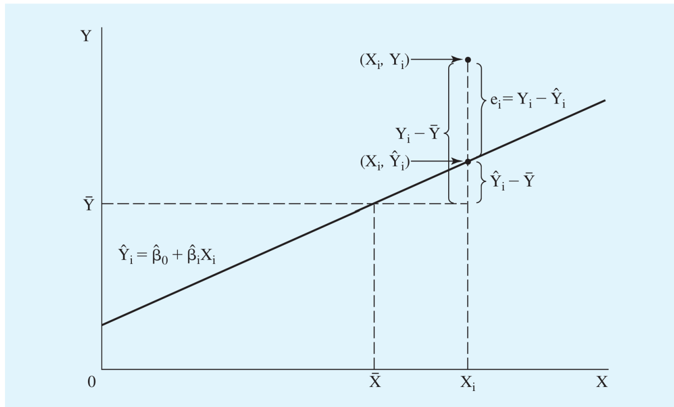
5. Note that the magnitude of the estimated coefficients is a function of the units in which the data are measured.

6. *Interpretation:* The estimated coefficients tell us to what degree and in what direction is the variation in the dependent variable explained by the variation in the independent variables.

   - Total Sum of Squares (TSS): $\sum (Y_i - \bar{Y})^2$
   - Explained Sum of Squares (ESS): $\sum (\hat{Y}_i - \bar{Y})^2$
   - Residual Sum of Squares (RSS): $\sum e_i^2$

7. Variance Decomposition: TSS = ESS + RSS

Figure 1: OLS Variance Decomposition. *Source:* Studenmund (2017)



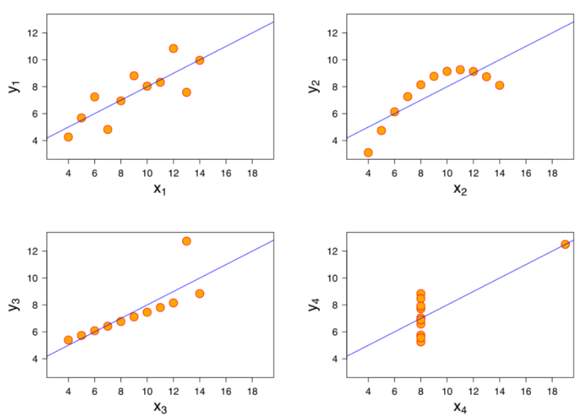
8. From the above, a good *ex-post* measure of *fit* is to measure how much of the total variation in the dependent variable has the regression explained. These measures have nothing to do with well-informed *ex-ante* decision making!

9. The Coefficient of Determination $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \in [0, 1]$

   The R^2 measures goodness-of-fit.

10. A note of caution #1: Anscombe's Quartet

Figure 2: Anscombe's Quartet. *Source:* Wikipedia



11. A note of caution #2: The ESS, which forms the numerator of the coefficient of determination, is of the form $\sum(\hat{\beta}_0 + \hat{\beta}_1 X_i - \bar{Y})^2$. This term is strictly increasing in the number of regressors. Dishonest researchers may be tempted to add in vacuous variables to push up the R^2 and validate the fit of their estimated regression! *Adding another independent variable **only** increases the* R^2.

This can lead to *spurious* regressions.

12. *Degrees of Freedom:* The amount of information allowed to vary freely. Think in terms of spans.

13. The *Adjusted* R^2 places a penalty for including more regressors and adjusts for *df*.

$$\bar{R}^2 = 1 - \frac{\sum e_i^2/(N-K-1)}{\sum(Y_i - \bar{Y})^2/(N-1)}$$

Here, N is the number of observations and K the number of regressors. Therefore, including the intercept, $K+1$ parameters are to be estimated.

Please attempt from the back of the chapter: Q1, Q2, Q3, Q5, Q6