Graduate Econometrics
The Linear Regression Model

Yonas Alem

Department of Economics
University of Gothenburg

November 5, 2014

- The linear regression model with one regressor is given by:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad (1)$$

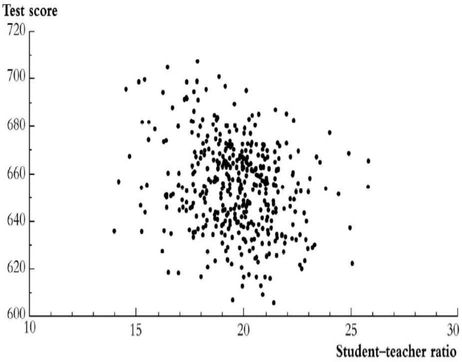
Where

- The subscript i runs over observations, $i = 1, ..., n$;
- Y_i is the dependent variable;
- X_i is the independent variable;
- $\beta_0 + \beta_1 X_i$ is the population regression line or the population regression function;
- β_0 is the intercept of the population regression line;
- β_1 is the slope of the population regression line; and
- u_i is the error term

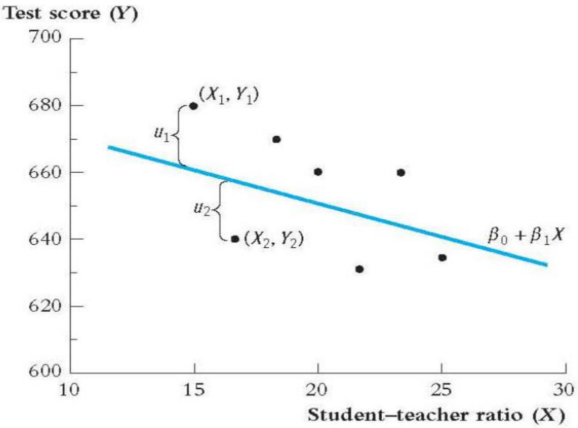
Linear Regression with One Regressor
Estimation with OLS (Ch. 4.2)

FIGURE 4.2 Scatterplot of Test Score vs. Student–Teacher Ratio (California School District Data)

Data from 420 California school districts. There is a weak negative relationship between the student–teacher ratio and test scores: The sample correlation is −0.23.



- Given a certain sample size, which linear combination, $\beta_0 + \beta_1 X_i$ gives a good approximation of Y_i?
- The Ordinary Least Square (OLS) estimator chooses the regression coefficients so that the estimated regression line is as close as possible to the observed data
- Let b_0 and b_1 be some estimators of β_0 and β_1. The regression line based these estimators is $b_0 + b_1 X$, so the value of Y_i predicted using this line is $b_0 + b_1 X$
- Thus, the mistake made in predicting the ith observation is $Y_i - (b_0 + b_1 X_i) = Y_i - b_0 - b_1 X_i$


Test score (Y)
700
680
660
640
620
600
10
15
20
25
30
Student–teacher ratio (X)
(X1, Y1)
u1
u2
(X2, Y2)
β0 + β1X

- The OLS estimators of $\hat{\beta_0}$ and $\hat{\beta_1}$ of the unknown parameters β_0 and β_1 are defined as those values of β_0 and β_1 that minimize the sum of the squared deviations given by

$$S(b) = \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2 \qquad (2)$$

- Where b_0 and b_1 denote any arbitrary values

- To minimize the sum of squared prediction mistakes given by $S(b)$, first take the partial derivatives with respect to b_0 and b_1

$$\frac{\partial S(b)}{\partial S(b_0}) = -2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i) \;\; and \tag{3}$$

$$\frac{\partial S(b)}{\partial S(b_1}) = -2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)X_i \;\; and \tag{4}$$

- Setting these two derivatives equal to zero gives the **normal equations:**

$$-2\sum(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \tag{5}$$

$$-2\sum X_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \tag{6}$$

- Rearranging terms, (5) $\Longrightarrow$:

$$\sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i \Longrightarrow \bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X} \Longrightarrow \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \tag{7}$$

- Rearranging (6):

$$\sum X_i Y_i = \hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2 \tag{8}$$

- Plugging (7) in (8):

$$\sum X_i Y_i = (\bar{Y} - \hat{\beta}_1 \bar{X}) \sum X_i + \hat{\beta}_1 \sum X_i^2 \tag{9}$$

$$\hat{\beta}_1 [\sum X_i^2 - \bar{X} \sum X_i] = \sum X_i Y_i - \bar{Y} \sum X_i \tag{10}$$

$$\hat{\beta}_1 [\sum X_i^2 - \frac{1}{n} (\sum X_i)^2] = \sum X_i Y_i - \frac{1}{n} \sum X_i \sum Y_i \tag{11}$$

- Rearranging:

$$\hat{\beta}_1 = \frac{\sum X_i Y_i - \frac{1}{n}\sum X_i \sum Y_i}{\sum X_i^2 - \frac{1}{n}(\sum X_i)^2} \tag{12}$$

- If you divide the numerator and denominator of (12) by $n-1$:

$$\hat{\beta}_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} \tag{13}$$

- (13) shows that $\hat{\beta}_1$ is the ratio of the sample covariance between X and Y and the sample variance of X, i.e.,

$$\hat{\beta}_1 = \frac{S_{XY}}{S_X^2} \tag{14}$$

The Linear Multiple Regression Model in Matrix Form (Ch. 18.1)

- The linear multiple regression model and the OLS estimator can each be represented compactly using matrix notation
- The population multiple regression model is given by:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + u_i, i = 1, ..., n. \quad (15)$$

- To write the multiple regression model in matrix form, define the following vectors:

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \boldsymbol{U} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{k1} \\ 1 & X_{12} & \cdots & X_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & \cdots & X_{kn} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1' \\ \boldsymbol{X}_2' \\ \vdots \\ \boldsymbol{X}_n' \end{pmatrix}, \quad (16)$$

The Linear Multiple Reg. Model in Matrix Form (Ch. 18.1) Cont.

and

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \tag{17}$$

- $\boldsymbol{Y}$ is the $n \times 1$ dimensional vector of n observations on the dependent variable
- $\boldsymbol{X}$ is the $n \times (k+1)$ dimensional matrix of n observations on the $k+1$ regressors (including the "constant" regressor for the intercept)
- The $(k+1) \times 1$ dimensional column vector $\boldsymbol{X_i}$ is the *ith* observation on the $k+1$ regressors; that is, $\boldsymbol{X}_i' = (1 \;\; X_{1i}...X_{ki})$, where $\boldsymbol{X_i}'$ denotes the transpose of $\boldsymbol{X_i}$

The Linear Multiple Reg. Model in Matrix Form (Ch. 18.1) Cont.

- $\boldsymbol{U}$ is the $n \times 1$ dimensional vector of n error terms
- $\boldsymbol{\beta}$ is the $(k+1) \times 1$ dimensional vector of $k+1$ unknown regression coefficients
- The multiple regression model in equation (15) for the i^{th} observation written using the vectors $\boldsymbol{\beta}$ and $\boldsymbol{X_i}$, is

$$Y_i = \mathbf{X}_i' \boldsymbol{\beta} + u_i, i = 1, ..., n. \quad (18)$$

- One can define the OLS estimators of b of the unknown parameters β to be the values of β that minimize the sum of squared deviations. If b denote any arbitrary $k+1$- element vector, then

$$S(b) = (Y - Xb)'(Y - Xb) = (Y' - (Xb)')(Y - Xb) \quad (19)$$

$$= (Y' - b'X')(Y - Xb) = (Y' - b'X')Y - (Y' - b'X')Xb \quad (20)$$

$$= Y'Y - b'X'Y - Y'Xb + b'X'Xb = Y'Y - 2Y'Xb + b'X'Xb \quad (21)$$

- Differentiate $S(b)$ w.r.t b to find the values of b that minimize $S(b)$

$$\frac{\partial S(b)}{\partial b} = -2X'Y + 2X'Xb = 0 \tag{22}$$

- Denote the value of b that minimizes $S(b)$, $\hat{\beta}$. Thus,

$$-2X'Y + 2X'X\hat{\beta} = 0 \tag{23}$$

$$X'X\hat{\beta} = X'Y \tag{24}$$

$$(X'X)^{-1}X'X\hat{\beta} = (X'X)^{-1}X'Y \Longrightarrow \tag{25}$$

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{26}$$

The Simple Linear Regression Model in Matrix Form

- The simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \tag{27}$$

- $Y = (n \times 1), X = (n \times 2), \beta(2 \times 1)$
- We saw that the OLS estimator of $\hat{\beta} = (X'X)^{-1}X'Y$. Thus

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = [\begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_{12} & X_{22} & \cdots & X_{n2} \end{pmatrix} \begin{pmatrix} 1 & X_{12} \\ 1 & X_{22} \\ \cdots & \cdots \\ 1 & X_{n2} \end{pmatrix}]^{-1}$$

$$\times \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_{12} & X_{22} & \cdots & X_{n2} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \tag{28}$$

- After simplifying:

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{pmatrix} n & \sum X_{i2} \\ \sum X_{i2} & \sum X_{i2}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum Y_i \\ \sum X_{i2} Y_i \end{pmatrix} \tag{29}$$

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \frac{1}{n \sum X_{i2}^2 - (\sum X_{i2})^2} \begin{pmatrix} \sum X_{i2}^2 & -\sum X_{i2} \\ -\sum X_{i2} & n \end{pmatrix} \begin{pmatrix} \sum Y_i \\ \sum X_{i2} Y_i \end{pmatrix} \tag{30}$$

- $\Longrightarrow$

$$\hat{\beta}_0 = \frac{\sum X_{i2}^2 \sum Y_i - \sum X_{i2} \sum X_{i2} Y_i}{n \sum X_{i2}^2 - (\sum X_{i2})^2} \tag{31}$$

$$\hat{\beta}_1 = \frac{-\sum X_{i2} \sum Y_i + n \sum X_{i2} Y_i}{n \sum X_{i2}^2 - (\sum X_{i2})^2} \tag{32}$$

The Simple Linear Regression Model in Matrix Form Cont.

- After rearranging:

$$\hat{\beta}_1 = \frac{\sum X_{i2} \sum Y_i - \frac{1}{n} \sum X_{i2} \sum Y_i}{\sum X_{i2}^2 - \frac{1}{n} (\sum X_{i2})^2} \tag{33}$$

- This is exactly the same as the formula for $\hat{\beta}_1$ which we derived in eq.(12) for the one regressor case without matrix algebra

- Consider the regression equation:

$$Y = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + u_i, i = 1, ..., n. \quad (34)$$

$$Y = X\beta + U \quad (35)$$

Where:

- $Y = (n \times 1)$ vector of dependent variables
- $X = n \times (K + 1)$ Matrix of explanatory variables
- $U = (n \times 1)$ Vector of random error terms
- $\beta = (k + 1) \times 1$ Vector of unknown population parameters

The Simple Linear Regression Model in Matrix Form Cont.
Assumptions (Ch. 4.4, 6.5, & 18.1)

1. $E(u_i|X_i) = 0$ or $E(U|X) = 0$;
2. $(X_i, Y_i), i = 1, ...n$, are independently and identically distributed (i.i.d) draws from their joint distribution $\Longrightarrow (X_i, u_i)$ are independent;
3. $X_i, \& u_i$ have nonzero finite fourth moments (large outliers are not likely);
4. X has full column rank (there is no perfect multicollinearity or the columns of X are linearly independent);
5. var $(u_i|X_i) = \sigma_u^2$ (homoskedasticity); and
6. The conditional distribution of u_i given X_i is normal (normal errors)

- Assumption (6) in addition to homoskedasticity
  $\Longrightarrow Cov\{u_i, u_j\} = 0 \; \forall i, j, \; i \neq j$

$$Cov\{U|X\} = \begin{pmatrix} E(u_1 u_1|X) & E(u_1 u_2|X) & \cdots E(u_1 u_n|X) \\ E(u_2 u_1|X) & E(u_2 u_2|X) & \cdots E(u_2 u_n|X) \\ \cdots & \cdots & \cdots \\ E(u_n u_1|X) & E(u_n u_2|X) & \cdots E(u_n u_n|X) \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_u^2 & 0 & \cdots 0 \\ 0 & \sigma_u^2 & \cdots 0 \\ \cdots & \cdots & \cdots \\ 0 & 0 & \cdots \sigma_u^2 \end{pmatrix}$$

$$= \sigma_u^2 I_n \quad (36)$$

- The OLS estimators $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$ are the values of $b_0, b_1, ..., b_k$ that minimize the sum of squared prediction mistakes $\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_{1i} - ... - b_k X_{ki})^2$. The **OLS predicted values** $\hat{Y}_i$ are

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + ... + \hat{\beta}_k X_{ki}, i = 1, ..., n, and \qquad (37)$$

- The **OLS residual** for the i^{th} observation is the difference between Y_i and its OLS predicted value; i.e., the OLS residual is:

$$\hat{u}_i = Y_i - \hat{Y}_i, i = 1, ..., n. \qquad (38)$$

- The OLS estimators $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$ and residual $\hat{u}_i$ are computed from a sample of n observations of $(X_{1i}, ..., X_{ki}, Y_i), i = 1, ..., n$. These are estimators of the unknown true population coefficients $\beta_0, \beta_1, ..., \beta_k$ and error term, u_i

- Three commonly used summary statistics in multiple regression:
  1. The standard error of the regression
  2. The regression R^2, and;
  3. The adjusted R^2 denoted $\bar{R}^2$
- All three statistics measure how well the OLS estimate of the multiple regression line describes, or "fits", the data

The Standard Error of the Reg.

- The standard error of the regression (SER) estimates the standard deviation of the error term u_i.
- It is thus a measure of the spread of the distribution of Y around the regression line.
- In a multiple regression, the SER is given by $SER = s_{\hat{u}}$

$$s_{\hat{u}}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}^2 = \frac{SSR}{n-k-1} \tag{39}$$

- and where $SSR = \sum_{i=1}^{n}\hat{u}^2$ is the sum of squared residuals, and k is the number of slope coefficients.
- In the case of a simple linear regression (i.e., only one regressor), the denominator reduces to $n-2$

The R^2

- The regression R^2 is the fraction of the sample variance of Y_i explained by (or predicted by) the regressors.
- Equivalently, the R^2 is 1 minus the fraction of the variance Y_i not explained by the regressors (ranges 0-1)

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS} \tag{40}$$

Where:

- ESS= Explained Sum of Squares $= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$ and the total sum of squares, $TSS = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$, and SSR = sum of squares of residuals (the fraction of the variance Y_i not explained by the regressors)
- In general SSR decreases as we add more explanatory variables and hence R^2 increases

The "Adjusted R^2"

- The R^2 increases when a new variable is added and an increase in R^2 does not mean that adding a variable actually improves the fit of the model
- In this case, R^2 gives an inflated estimate of how well the regression fits the data
- The **adjusted** R^2 (as the name implies) corrects or adjusts for this
- The adjusted R^2 is a modified version of the R^2 that does not necessarily increase when a new regressor is added

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS} = 1 - \frac{S_{\hat{u}^2}}{S_Y^2} \qquad (41)$$

The "Adjusted R^2" cont.

- Three useful things to know about the $\bar{R}^2$
  1. $(n-1/n-k-1)$ is always greater than 1, so $\bar{R}^2$ is always less than R^2
  2. Adding a regressor has two opposite effects on the $\bar{R}^2$: the SSR falls (and this increases the $\bar{R}^2$). On the other hand, the factor $(n-1/n-k-1)$ increases. Whether the $(n-1/n-k-1)$ increases or decreases depends on which of the two effects is stronger
  3. The $\bar{R}^2$ can be negative. This happens when the regressors, taken together, reduce the SSR by such a small amount that this reduction fails to offset the factor $(n-1/n-k-1)$.

Interpretation of R^2 & $\bar{R}^2$

- Four potential pitfalls to guard against when using R^2 or $\bar{R}^2$
  1. An increase in the R^2 or $\bar{R}^2$ does not mean that an added variable is statistically significant. (statistical significance requires hypothesis testing)
  2. A high R^2 or $\bar{R}^2$ does not mean that the regressors are a true cause of the dependent variable (spurious regression)
  3. A high R^2 or $\bar{R}^2$ does not mean that there is no omitted variable bias
  4. A high R^2 or $\bar{R}^2$ does not mean that you have the most appropriate set of regressors, not does a low A R^2 or $\bar{R}^2$ does not mean that necessarily mean that you have an inappropriate set of regressors

- For the multiple regression equation given in a matrix form

$$Y = \boldsymbol{X\beta} + U, i = 1, ..., n. \quad (42)$$

- We showed that:

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (43)$$

- $E\{\hat{\beta}\} = E\{\beta + (X'X)^{-1}X'U\} = \beta + (X'X)^{-1}X'E\{U\} = \beta$
- Because $E\{U\} = 0$ by assumption (1)
- Hence, $\hat{\beta}$ is an unbiased estimator of β!

Covariance of $\hat{\beta}$ cont.

- $cov\{\hat{\beta}\} = E\{(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\}$
- We can show that:
  $(\hat{\beta} - \beta) = (X'X)^{-1}X'(X\beta + U) - \beta = (X'X)^{-1}X'U \Longrightarrow$

$$cov\{\hat{\beta}\} = E\{(X'X)^{-1})X'UU'X(X'X)^{-1}\}$$

$$= (X'X)^{-1}X'E\{UU'\}X(X'X)^{-1}$$

$$= (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1}$$

$$= \sigma^2(X'X)^{-1}$$

Covariance of $\hat{\beta}$

- $cov\{\hat{\beta}\}$ is therefore a $(k+1 \times k+1)$ matrix showing
  - The Variances in the main diagonal
  - The covariances in the off-diagonal
- The population variance σ^2 is however not known. The reasonable option is to estimate it from the sample variance of the residuals, which we presented in eq. (39)
- In matrix form:

$$\hat{u} = Y - X\hat{\beta}$$

$$\hat{S_u}^2 = \frac{1}{(n-k-1)} \sum_{i=1}^{n} \hat{u_i}^2 = \frac{\hat{u_i}' \hat{u_i}}{(n-k-1)} \qquad (44)$$

- Thus,

$$cov(\hat{\beta}) = \hat{S}^2_{\hat{u}} (X'X)^{-1}$$

- If four of the least square assumptions $1, 2, 3\&5$ hold, then OLS estimator has the smallest variance, conditional on $X_1, ..., X_n$ among all estimators in the class of linear conditionally unbiased estimators.
  - I.e., the OLS estimator is Best Linear Conditionally Unbiased Estimator **(BLUE)**
- Limitations of the Guass-Markov theorem:
  - The conditions might not hole in practice (e.g., if the errors are heteroskedastic). In this case, OLS is not best and we may have to use other estimators (e.g., weighted least squares)
  - Even if the conditions of the theorem hold, there are other candidate estimators that are not linear and conditionally unbiased; under some conditions, these other estimators are more efficient than OLS