Topic 4: The Multiple Linear Regression Model

The Multiple Linear Regression Model (MLRM)

- The starting point for the bivariate linear regression model (BLRM) is the assumption that

$$E(y_i|x_i) = \beta_0 + \beta_1 x_i, i = 1, 2, ..., n. \tag{1}$$

- In (1) the conditional mean of y is a linear function of a single explanatory variable, x.
- The multiple linear regression model (MLRM) generalizes the BLRM by allowing the conditional mean of y to contain k explanatory variable, where k is an arbitrary scalar.
- In the multiple linear regression model we postulate that

$$E(y_i|x_{i1}, x_{i2}, ..., x_{ik}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik}, \; i = 1, 2, ...n, \tag{2}$$

where x_{ij} denotes the ith observation on the explanatory variable x_j.

- Note that since we now have k explanatory variables in the model for the conditional mean of y, we need two subscripts for each x. The first subscript identifies the observation and the second subscript identifies the explanatory variable.
- For example, the notation x_{12} denotes the first sample observation on the explanatory variable x_2.
- The BLRM is the special case of (2) obtained by setting

$$k = 1.$$

- As in the bivariate model, denote the deviation of y_i from its conditional mean by

$$\begin{aligned} u_i &= y_i - E(y_i | x_{i1}, x_{i2}, ..., x_{ik}) \\ &= y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - ..... - \beta_k x_{ik}, \; i = 1, 2, ...n. \end{aligned} \tag{3}$$

- Rearranging (3) we obtain the general form of the multiple linear regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + u_i, \;\; i = 1, 2, ...n. \qquad (4)$$

- Note that since

$$\begin{aligned} E(y_i | x_{i1}, x_{i2}, ..., x_{ik}) &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik}, \\ E(y_i | x_{i1} + 1, x_{i2}, ..., x_{ik}) &= \beta_0 + \beta_1 (x_{i1} + 1) + \beta_2 x_{i2} + ... + \beta_k x_{ik}, \end{aligned}$$

if follows that

$$E(y_i | x_{i1} + 1, x_{i2}, ..., x_{ik}) - E(y_i | x_{i1}, x_{i2}, ..., x_{ik}) = \beta_1. \qquad (5)$$

- Equation (5) states that the regression coefficient β_1 in

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + u_i, \;\; i = 1, 2, ...n, \qquad (4)$$

  measures the change in $E(y_i | x_{i1}, x_{i2}, ..., x_{ik})$ associated with a one unit change in x_{i1}, holding fixed the values of the explanatory variables $(x_{i2}, ..., x_{ik})$.

- In general, the regression coefficient β_k in

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + u_i, \;\; i = 1, 2, ...n, \qquad (4)$$

  measures the change in $E(y_i | x_{i1}, x_{i2}, ..., x_{ik})$ associated with a one unit change in x_{ik}, holding fixed the values of the other $k - 1$ explanatory variables.

- Again, using the rules of elementary calculus,

$$\frac{\partial E(y_i|x_{i1}+1, x_{i2}, ..., x_{ik})}{\partial x_{ik}} = \beta_k. \tag{6}$$

- We interpret (6) as stating that regression coefficient β_k measures the change in $E(y_i|x_{i1}, x_{i2}, ..., x_{ik})$ associated with a small change in x_{ik}, holding fixed the values of the other $k-1$ explanatory variables.
- When (4) is estimated by OLS the predicted, or estimated, value of y_i is given by

$$\widehat{y}_i = \widehat{E}(y_i|x_{i1}, x_{i2}, ..., x_{ik}) = \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \widehat{\beta}_2 x_{i2} + ... + \widehat{\beta}_k x_{ik},$$

where $\widehat{\beta}_j$, $j = 0, 1, 2, ..., k$, denotes the OLS estimate of β_j.

- The *ith* OLS residual is defined as

$$\begin{aligned} \widehat{u}_i &= y_i - \widehat{y}_i \\ &= y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_{i1} - \widehat{\beta}_2 x_{i2} - ..... - \widehat{\beta}_k x_{ik}, \end{aligned} \quad (7)$$

and

$$\begin{aligned} SSR &= \sum_{i=1}^{n} \widehat{u}_i^2 \\ &= \sum_{i=1}^{n} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_{i1} - \widehat{\beta}_2 x_{i2} - ..... - \widehat{\beta}_k x_{ik} \right)^2 . \end{aligned} \quad (8)$$

- As in the BLRM,

$$R^2 = \frac{SSE}{SST} = \frac{\sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{9}$$

and, as long as we have an intercept in the MLRM,

$$0 \leq R^2 \leq 1. \tag{10}$$

- In the MLRM we interpret R^2 is a measure of the proportion of the variation in the dependent variable in our sample (**the sample variation**) that is "explained" or "predicted" by the explanatory variables $x_1, .x_2, ..., x_k$.

- Recall that in Topic 3 we estimated a BLRM of the relationship between weekly wages and education given by

$$wage_i = \beta_0 + \beta_1 educ_i + u_i, i = 1, 2, ...935. \tag{11}$$

- When (11) is estimated by OLS we obtain the results reported in Figure 1 below.

Dependent Variable: WAGE
Method: Least Squares
Sample: 1 935
Included observations: 935

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	146.9524	77.71496	1.890916	0.0589
EDUC	60.21428	5.694982	10.57322	0.0000

R-squared	0.107000	Mean dependent var	957.9455
Adjusted R-squared	0.106043	S.D. dependent var	404.3608
S.E. of regression	382.3203	Akaike info criterion	14.73253
Sum squared resid	1.36E+08	Schwarz criterion	14.74289

Figure: 1

- Based on the output reported in Figure 1 we can deduce the following conclusions:

- $$\widehat{wage}_i = \widehat{E}(wage_i|educ_i) = 146.95 + 60.21 educ_i.$$

  For example, the predicted average weekly wage of the population of individuals with 10 years of education is

  $$\begin{aligned} \widehat{wage}_i &= \widehat{E}(wage_i|educ_i = 10) = 146.95 + 60.21(10) \\ &= \$749.05. \end{aligned}$$

- Since

  $$\widehat{\beta}_1 = 60.21,$$

  we predict that an extra year of education increases average weekly wages by $60.21.
- Since

  $$R^2 = 0.107,$$

  we conclude that variation in years of education explains 10.7 % the variation in average weekly wages in our sample.

- In Topic 3 we posed the question of whether or not we could interpret $\widehat{\beta}_1$ as the estimated **causal effect** of education on average weekly wages?
- That is, can we conclude that an extra year of education is predicted to **cause** average weekly wages to increase by \$60.21?
- In the present example, the treatment is years of education and different individuals in our sample **have chosen** to receive different levels of the treatment.
- We must ask ourselves whether or not there are systematic differences between the individuals who have chosen different levels of the treatment (years of education) and if there are, do these differences affect the dependent variable (wages).
- Common sense would suggest that:

  - On average, people who choose to obtain more of education are more intelligent than those who choose to obtain fewer years of education.
  - More intelligent people are likely to be more productive in the workplace and receive higher wages as a consequence.
- These observations suggest that in order to estimate the causal effect of education on wages, we need to estimate the average wage of those people who have different levels of education, but the same level of intelligence.
- If we control for intelligence in this way, then we can plausibly attribute observed differences in average wages between two groups with different levels of education, to differences in their levels of education.
- In addition to data on wages and years of education, we also have data on the IQ scores of the individuals in our sample.

- We next include IQ score as an additional explanatory variable in our model of wage determination and estimate the MLRM given by

$$wage_i = \beta_0 + \beta_1 educ_i + \beta_2 IQ_i + u_i, i = 1, 2, ...935. \tag{12}$$

- When we estimate (12) by OLS we obtain the results reported in Figure 2 below.

Dependent Variable: WAGE
Method: Least Squares
Sample: 1 935
Included observations: 935

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-128.8899	92.18232	-1.398206	0.1624
EDUC	42.05762	6.549836	6.421171	0.0000
IQ	5.137958	0.955827	5.375403	0.0000

R-squared	0.133853	Mean dependent var	957.9455
Adjusted R-squared	0.131995	S.D. dependent var	404.3608
S.E. of regression	376.7300	Akaike info criterion	14.70414
Sum squared resid	1.32E+08	Schwarz criterion	14.71967

Figure: 2

- Based on the output reported in Figure 2 we can deduce the following conclusions:

- $$\widehat{wage}_i = \widehat{E}(wage_i | educ_i, IQ_i) = -128.89 + 42.06 educ_i + 5.14 IQ_i.$$

  For example, we predict that the average weekly wage for the subset of the population with 10 years of education and an IQ score of 100 is

  $$\begin{aligned} \widehat{wage}_i &= E(wage_i | 10, 100) = -128.89 + 42.06(10) + 5.14(100) \\ &= \$805.71. \end{aligned}$$

- Since
  $$\widehat{\beta}_1 = 42.06,$$
  we predict that an extra year of education increases the average weekly wage of the population by \$42.06, **controlling for IQ score.**

- That is, we predict that the average weekly wage of the subset of the population with a given IQ score and x+1 years of education, will exceed by $42.06 the average weekly wage of the subset of the population with the same IQ score and x years of education.
- Because we are controlling for differences in intelligence by including IQ score as a regressor in the regression equation, we can plausibly argue that the $\widehat{\beta}_1$ we obtain when we estimate

$$wage_i = \beta_0 + \beta_1 educ_i + \beta_2 IQ_i + u_i, i = 1, 2, ...935, \tag{12}$$

  is an estimate of the **causal effect** of education of the average wage.
- Of course, not everyone would agree with this statement. Some observers would argue that IQ score is a inadequate measure of intelligence and that including it in the regression equations does not really control for intelligence.
- This is something that well trained, intelligent people can disagree about!

- It is important to realize that the interpretation of β_1 in (12) is different from the interpretation of β_1 in

$$wage_i = \beta_0 + \beta_1 educ_i + u_i, i = 1, 2, ...935, \tag{11}$$

  - In (12) β_1 measures the marginal effect of education on average wages, **controlling for the influence of IQ score on average wages**.
  - In (11) β_1 measures the marginal effect of education on wages without controlling for the influence of any other variables on average wages.
  - The fact that we control for the influence of IQ score (intelligence) in (12) is what allows us to plausibly interpret β_1 as the **causal effect** of education on average wages.

- It is instructive to compare the estimated marginal effect of education on average weekly wages from estimating "**the short regression**"

$$wage_i = \beta_0 + \beta_1 educ_i + u_i, i = 1, 2, ...935, \tag{11}$$

and the "**long regression**" given by

$$wage_i = \beta_0 + \beta_1 educ_i + \beta_2 IQ_i + u_i, i = 1, 2, ...935. \tag{12}$$

- When we estimate (11),

$$\widehat{\beta}_1 = 60.21$$

and when we estimate (12)

$$\widehat{\beta}_1 = 42.06.$$

- Comparing these numbers suggests that part of the effect on average weekly wages attributed to education when we estimate (11) is really due to differences in IQ score (or intelligence) across individuals.
- Consequently,
$$\widehat{\beta}_1 = 60.21$$
is an overestimate of the effect of education on average weekly wages.
- Because education and IQ score are positively correlated, an increase in individual education is associated with an increase in individual intelligence, and when we estimate (11) we cannot identify the separate effects of these two factors.
- A key lesson to be learned from this simple example is that one must think very carefully about what explanatory variables one needs to include in a regression equation if one wishes to give a causal interpretation to a particular regression coefficient.

- We are very fortunate to have data on individual IQ score in this data set. One rarely has information on such a personal characteristic.
- This poses the question, "could we estimate the causal effect of education on averages wages if we did not have data on IQ score, or some similar proxy for intelligence"?
- It turns out that sometimes **we may** be able to estimate the causal effect of education on average wages by using an alternative estimation procedure to OLS called **instrumental variables (IV) estimation**.
- However, IV estimation is beyond the scope of this unit.
- The curious student will have to enrol in ETC3410 to find out about IV estimation!

- In order to derive the formula for the OLS estimator of the regression coefficients in the MLRM given by

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + u_i, \;\; i = 1, 2, ...n, \tag{4}$$

  and to analyze the statistical properties of the OLS estimator, it is convenient to formulate the model in the language of matrix algebra.
- Writing out the n equations in (4) we obtain

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + ... + \beta_k x_{1k} + u_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + ... + \beta_k x_{2k} + u_2 \\ &. \\ &. \\ y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + ... + \beta_k x_{nk} + u_n \end{aligned} \tag{13}$$

- Collecting the n observations on y into an nx1 vector, and the n observations on each of the $k+1$ regressors, $(x_{1,}x_2, ..., x_k)$, into an nxk matrix, (13) may be written as

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & . & . & x_{1k} \\ 1 & x_{21} & x_{22} & . & . & x_{2k} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & x_{n1} & x_{n2} & . & . & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ u_n \end{bmatrix}, \quad (14)$$

or, more compactly, as

$$\underset{(nx1)}{y} = \underset{(nxk+1)}{X} \underset{(k+1x1)}{\beta} + \underset{(nx1)}{u}, \quad (15)$$

where

$$\underset{(nx1)}{y} = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{bmatrix}, \underset{(k+1x1)}{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ \beta_k \end{bmatrix}, \underset{(nx1)}{u} = \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ u_n \end{bmatrix},$$

and

$$\underset{nx(k+1)}{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & . & . & x_{1k} \\ 1 & x_{21} & x_{22} & . & . & x_{2k} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & x_{n1} & x_{n2} & . & . & x_{nk} \end{bmatrix}. \quad (16)$$

- Notice that the conformability conditions for both matrix multiplication and matrix addition are satisfied in the matrix equation

$$\underset{(nx1)}{y} = \underset{(nxk+1)}{X} \underset{(k+1x1)}{\beta} + \underset{(nx1)}{u}. \tag{15}$$

- The first column of the X matrix in (16) is a vector of 1s (to capture the intercept) and each of the remaining columns contains n observations on a particular regressor.
- Let

$$\widehat{\beta} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ . \\ . \\ \widehat{\beta}_k \end{bmatrix} \tag{17}$$

denote the OLS estimator of β we obtain when we estimate (15) by OLS.

- The $nx1$ vector of predicted values of y is given by

$$\widehat{y} = X\widehat{\beta}, \tag{18}$$

and the $nx1$ vector of OLS residuals is given by

$$\begin{aligned} \widehat{u} &= y - \widehat{y} \\ &= y - X\widehat{\beta}. \end{aligned} \tag{19}$$

S1 Rearranging (19) we obtain

$$y = X\widehat{\beta} + \widehat{u}. \tag{20}$$

S2 Multiplying on both sides of (20) by X' we obtain

$$X'y = X'X\widehat{\beta} + X'\widehat{u}. \tag{21}$$

S3 It can be shown that in the MLRM

$$\underset{(k+1xn)}{X'}\underset{(nx1)}{\widehat{u}} = \underset{(k+1x1)}{\mathbf{0}}, \tag{22}$$

(see Appendix), so (21) reduces to

$$X'y = X'X\widehat{\beta}. \tag{23}$$

S4 Multiplying on both sides of (23) by the matrix $(X'X)^{-1}$ we obtain

$$\begin{aligned}(X'X)^{-1}X'y &= (X'X)^{-1}X'X\widehat{\beta} \\ &= I_{k+1}\widehat{\beta} \\ &= \widehat{\beta}.\end{aligned}$$

Therefore, the OLS estimator of β in the MLRM

$$y = X\beta + u, \tag{24}$$

is given by the formula

$$\widehat{\beta} = (X'X)^{-1}X'y. \tag{25}$$

(This is the most famous formula in econometrics!).

- At S4 in the derivation we implicitly assumed that the matrix $(X'X)^{-1}$ exists. That is, we assumed that $(X'X)$ is a nonsingular matrix.
- It can be shown that as long as the columns of X are linearly independent, this is a valid assumption to make.
- If any column of X can be expressed as a linear function of one or more of the remaining columns, the matrix $(X'X)$ is a **singular matrix** and the OLS estimator is not defined.
- When $(X'X)$ is a singular matrix, the regression model is said to suffer from exact, or perfect, **multicollinearity**.
- Perfect multicollinearity rarely occurs and, when it does, it is usually caused by an error in the choice of explanatory variables included in the regression equation.
- We will discuss the phenomenon of multicollinearity in more detail later in the unit.

- In this appendix we prove that

$$X'\widehat{u} = \mathbf{0},$$

a result that we used in our derivation of the OLS estimator in the MLRM.

- If we partition X into its constituent columns we obtain

$$\underset{(nxk+1)}{X} = \left[\underset{(nx1)}{i} \quad \underset{(nx1)}{x_1} \quad \underset{(nx1)}{x_2} \quad . \quad . \quad \underset{(nx1)}{x_k} \right], \tag{A1.1}$$

implying that

$$\underset{(k+1xn)}{X'} = \begin{bmatrix} \underset{(1xn)}{i'} \\ \underset{(1xn)}{x_1'} \\ \underset{(1xn)}{x_2'} \\ . \\ . \\ x_k' \end{bmatrix}.$$

- Therefore,

$$\underset{(k+1xn)}{X'}\underset{(nx1)}{\widehat{u}} = \begin{bmatrix} \underset{(1xn)}{i'} \\ \underset{(1xn)}{x_1'} \\ \underset{(1xn)}{x_2'} \\ . \\ . \\ \underset{(1xn)}{x_k'} \end{bmatrix} \underset{(nx1)}{\widehat{u}} = \begin{bmatrix} \underset{(1x1)}{i'\widehat{u}} \\ \underset{(1x1)}{x_1'\widehat{u}} \\ \underset{(1x1)}{x_2'\widehat{u}} \\ . \\ . \\ \underset{(1x1)}{x_k'\widehat{u}} \end{bmatrix}. \qquad \text{(A1.2)}$$

- Recall that in Topic 3 we saw that in the BLRM

$$y_i = \beta_0 + \beta_1 x_i + u_i . i = 1, 2, ...n,$$

the OLS residual vector is orthogonal to the vector x. That is,

$$x'\widehat{u} = 0.$$

Appendix III

- It can be shown that this result from the BLRM generalizes to the MLRM, in the sense that the OLS residual vector is orthogonal to each of the regressors. That is,

$$i'\widehat{u} = 0, x_1'\widehat{u} = 0, x_2'\widehat{u} = 0, ..., x_k'\widehat{u} = 0 \tag{A1.3}$$

- Substituting A1.3 into A1.2 we obtain

$$X'\widehat{u} = \begin{bmatrix} i'\widehat{u} \\ x_1'\widehat{u} \\ x_2'\widehat{u} \\ . \\ . \\ x_k'\widehat{u} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ . \\ . \\ 0 \end{bmatrix}. \tag{A1.4}$$