Introductory Econometrics

Lecture 10: Multiple regression model

Instructor: Ma, Jun

Renmin University of China

October 14, 2021

Why we need a multiple regression model

- There are many factors affecting the outcome variable Y.
- If we want to estimate the marginal effect of one of the factors (regressors), we need to control for other factors.
- Suppose that we are interested in the effect of X_1 on Y, but Y is affected by both X_1 and X_2:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + U_i.$$

- Suppose we regress Y only against X_1:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \left(X_{1,i} - \bar{X}_1\right) Y_i}{\sum_{i=1}^{n} \left(X_{1,i} - \bar{X}_1\right)^2}.$$

Omitted variable bias

Since Y depends on X_2 : $Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + U_i$,

- We have:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n \left(X_{1,i} - \bar{X}_1\right)\left(\beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + U_i\right)}{\sum_{i=1}^n \left(X_{1,i} - \bar{X}_1\right)^2} \\
&= \beta_1 + \beta_2 \frac{\sum_{i=1}^n \left(X_{1,i} - \bar{X}_1\right) X_{2,i}}{\sum_{i=1}^n \left(X_{1,i} - \bar{X}_1\right)^2} + \frac{\sum_{i=1}^n \left(X_{1,i} - \bar{X}_1\right) U_i}{\sum_{i=1}^n \left(X_{1,i} - \bar{X}_1\right)^2}.
\end{aligned}$$

- Assume that $\mathrm{E}\left[U_i \mid X_{1,i}, X_{2,i}\right] = 0$. Now, conditional on X's:

$$\mathrm{E}\left[\hat{\beta}_1\right] = \beta_1 + \beta_2 \frac{\sum_{i=1}^n \left(X_{1,i} - \bar{X}_1\right) X_{2,i}}{\sum_{i=1}^n \left(X_{1,i} - \bar{X}_1\right)^2} \neq \beta_1.$$

  The exception is when

$$\sum_{i=1}^n \left(X_{1,i} - \bar{X}_1\right) X_{2,i} = 0.$$

Omitted variable bias

- When the true model is

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + U_i,$$

  but we regress only on X_1,

$$Y_i = \beta_0 + \beta_1 X_{1,i} + V_i,$$

  where V_i is the new error term:

$$V_i = \beta_2 X_{2,i} + U_i.$$

- If X_1 and X_2 are related, we can no longer say that $\mathrm{E}\left[V_i \mid X_{1,i}\right] = 0$.
- When X_1 changes, X_2 changes as well, which contaminates estimation of the effect of X_1 on Y.
- As a result, $\hat{\beta}_1$ from the regression of Y on X_1 alone is biased.

Multiple linear regression model

- The econometrician observes the data:

$$\left\{ \left(Y_i, X_{1,i}, X_{2,i}, \ldots, X_{k,i} \right) : i = 1, \ldots, n \right\}.$$

- The model:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_k X_{k,i} + U_i,$$
$$\mathrm{E}\left[U_i \mid X_{1,i}, X_{2,i}, \ldots, X_{k,i} \right] = 0.$$

- We also assume no multicollinearity: None of the regressors are constant and there are no exact linear relationships among the regressors.

Interpretation of the coefficients

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_k X_{k,i} + U_i.$$

- β_j is a partial (marginal) effect of X_j on Y:

$$\beta_j = \frac{\partial Y_i}{\partial X_{j,i}}.$$

- For example, β_1 is the effect of X_1 on Y while holding the other regressors constant (or controlling for $X_2, \ldots, X_k$)

$$\Delta Y = \beta_1 \Delta X_1.$$

- In data, the values of all regressors usually change from observation to observation. If we do not control for other factors, we cannot identify the effect of X_1.

Changing more than one regressor simultaneously

- There are cases when we want to change more than one regressor at the same time to find an effect on Y.
- Example 3.2: the results from 526 observations on workers

$$\widehat{\log Wage} = 0.284 + 0.92 edu + 0.0041 exper + 0.22 tenure.$$

- The effect of staying one more year at the same firm: increasing both exper and tenure.
- Holding edu fixed,

$$\widehat{\log Wage} = 0.0041 \Delta exper + 0.22 \Delta tenure.$$

Modelling nonlinear effects

- Recall that in $Y_i = \beta_0 + \beta_1 X_i + U_i$, the effect of X_i on Y_i is linear: $dY_i/dX_i = \beta_1$ and constant for all values of X_i.
  Multiple regression can be used to model nonlinear effects of regressors.
- To model nonlinear returns to education, consider the following equation:

$$\log \text{Wage}_i = \beta_0 + \beta_1 \text{Education}_i + \beta_2 \text{Education}_i^2 + U_i,$$

  were Education_i = years of education of individual i.
- In this case, the return to education is:

$$\frac{d \log \text{Wage}_i}{d\text{Education}_i} = \beta_1 + 2\beta_2 \text{Education}_i.$$

  Now, return to education depends on years of education. For example, diminishing returns to education correspond to $\beta_2 < 0$.

OLS estimation

- The OLS estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ are the values that minimize the squared errors function:

$$\min_{b_0, b_1, \ldots, b_k} Q_n\left(b_0, b_1, \ldots, b_k\right), \text{ where}$$

$$Q_n\left(b_0, b_1, \ldots, b_k\right) = \sum_{i=1}^{n}\left(Y_i - b_0 - b_1 X_{1,i} - \ldots - b_k X_{k,i}\right)^2.$$

- The partial derivative with respect to b_0 is

$$\frac{\partial Q_n\left(b_0, b_1, \ldots, b_k\right)}{\partial b_0} = -2\sum_{i=1}^{n}\left(Y_i - b_0 - b_1 X_{1,i} - \ldots - b_k X_{k,i}\right).$$

- The partial derivative with respect to b_j, $j = 1, \ldots, k$ is

$$\frac{\partial Q_n\left(b_0, b_1, \ldots, b_k\right)}{\partial b_j} = -2\sum_{i=1}^{n}\left(Y_i - b_0 - b_1 X_{1,i} - \ldots - b_k X_{k,i}\right) X_{j,i}.$$

Normal equations (first-order conditions for OLS)

The OLS estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ are obtained by solving the following system of normal equations:

$$
\begin{aligned}
\sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - \ldots - \hat{\beta}_k X_{k,i}\right) &= 0, \\
\sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - \ldots - \hat{\beta}_k X_{k,i}\right) X_{1,i} &= 0, \\
\vdots &= \vdots \\
\sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - \ldots - \hat{\beta}_k X_{k,i}\right) X_{k,i} &= 0.
\end{aligned}
$$

Normal equations (first-order conditions for OLS)

- Since the fitted residuals are

$$\hat{U}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - \ldots - \hat{\beta}_k X_{k,i},$$

  the normal equations can be written as

$$\begin{aligned} \sum_{i=1}^{n} \hat{U}_i &= 0, \\ \sum_{i=1}^{n} \hat{U}_i X_{1,i} &= 0, \\ \vdots &= \vdots \\ \sum_{i=1}^{n} \hat{U}_i X_{k,i} &= 0. \end{aligned}$$

- We choose $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ so that $\hat{U}$'s and regressors are orthogonal (uncorrelated in sample).

Partitioned regression

- A representation for individual $\hat{\beta}$'s can be obtained through the partitioned regression result. Suppose we want to obtain an expression for $\hat{\beta}_1$.
- Consider first regressing $X_{1,i}$ against other regressors and a constant:

$$X_{1,i} = \hat{\gamma}_0 + \hat{\gamma}_2 X_{2,i} + \ldots + \hat{\gamma}_k X_{k,i} + \tilde{X}_{1,i},$$

  where $\hat{\gamma}_0, \hat{\gamma}_2, \ldots, \hat{\gamma}_k$ are the OLS coefficients, and $\tilde{X}_{1,i}$ is the fitted OLS residual:

$$\sum_{i=1}^{n} \tilde{X}_{1,i} = 0, \text{ and } \sum_{i=1}^{n} \tilde{X}_{1,i} X_{j,i} = 0 \text{ for } j = 2, \ldots, k.$$

- Then $\hat{\beta}_1$ can be written as

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \tilde{X}_{1,i} Y_i}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2}.$$

Proof of the partitioned regression result

- We can write $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1,i} + \hat{\beta}_2 X_{2,i} + \ldots + \hat{\beta}_k X_{k,i} + \hat{U}_i$, where $\sum_{i=1}^n \hat{U}_i = \sum_{i=1}^n \hat{U}_i X_{1,i} = \ldots = \sum_{i=1}^n \hat{U}_i X_{k,i} = 0$.
- Now,

$$\frac{\sum_{i=1}^n \tilde{X}_{1,i} Y_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2} =$$

$$\frac{\sum_{i=1}^n \tilde{X}_{1,i} \left(\hat{\beta}_0 + \hat{\beta}_1 X_{1,i} + \hat{\beta}_2 X_{2,i} + \ldots + \hat{\beta}_k X_{k,i} + \hat{U}_i\right)}{\sum_{i=1}^n \tilde{X}_{1,i}^2}$$

$$= \hat{\beta}_0 \frac{\sum_{i=1}^n \tilde{X}_{1,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} + \hat{\beta}_1 \frac{\sum_{i=1}^n \tilde{X}_{1,i} X_{1,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} +$$

$$+ \hat{\beta}_2 \frac{\sum_{i=1}^n \tilde{X}_{1,i} X_{2,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} + \ldots + \hat{\beta}_k \frac{\sum_{i=1}^n \tilde{X}_{1,i} X_{k,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} + \frac{\sum_{i=1}^n \tilde{X}_{1,i} \hat{U}_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2}.$$

Proof of the partitioned regression result

$$\frac{\sum_{i=1}^n \tilde{X}_{1,i} Y_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2} = \hat{\beta}_0 \frac{\sum_{i=1}^n \tilde{X}_{1,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} + \hat{\beta}_1 \frac{\sum_{i=1}^n \tilde{X}_{1,i} X_{1,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} + \\ + \hat{\beta}_2 \frac{\sum_{i=1}^n \tilde{X}_{1,i} X_{2,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} + \ldots + \hat{\beta}_k \frac{\sum_{i=1}^n \tilde{X}_{1,i} X_{k,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} + \frac{\sum_{i=1}^n \tilde{X}_{1,i} \hat{U}_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2}.$$

We will show that:

1. $\sum_{i=1}^n \tilde{X}_{1,i} = 0$.
2. $\sum_{i=1}^n \tilde{X}_{1,i} X_{2,i} = \ldots = \sum_{i=1}^n \tilde{X}_{1,i} X_{k,i} = 0$.
3. $\sum_{i=1}^n \tilde{X}_{1,i} X_{1,i} = \sum_{i=1}^n \tilde{X}_{1,i}^2$.
4. $\sum_{i=1}^n \tilde{X}_{1,i} \hat{U}_i = 0$.

Then

$$\frac{\sum_{i=1}^n \tilde{X}_{1,i} Y_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2} = \hat{\beta}_1.$$

Proof of the partitioned regression result (steps 1-2)

- $\tilde{X}_{1,i}$ is the fitted OLS residual:

$$X_{1,i} = \hat{\gamma}_0 + \hat{\gamma}_2 X_{2,i} + \ldots + \hat{\gamma}_k X_{k,i} + \tilde{X}_{1,i},$$

  where $\hat{\gamma}_0, \hat{\gamma}_2, \ldots, \hat{\gamma}_k$ are the OLS coefficients.
- The normal equations for this regression are:

$$\begin{aligned}
\sum_{i=1}^{n} \tilde{X}_{1,i} &= 0, \\
\sum_{i=1}^{n} \tilde{X}_{1,i} X_{2,i} &= 0, \\
\vdots &= \vdots \\
\sum_{i=1}^{n} \tilde{X}_{1,i} X_{k,i} &= 0.
\end{aligned}$$

Proof of the partitioned regression result (step 3)

Again, because $\tilde{X}_{1,i}$ are the OLS residuals (fitted) from the regression of X_1 against $X_2, \ldots, X_k$:

$$
\begin{aligned}
\sum_{i=1}^{n} \tilde{X}_{1,i} X_{1,i} &= \sum_{i=1}^{n} \tilde{X}_{1,i} \left(\hat{\gamma}_0 + \hat{\gamma}_2 X_{2,i} + \ldots + \hat{\gamma}_k X_{k,i} + \tilde{X}_{1,i} \right) \\
&= \hat{\gamma}_0 \sum_{i=1}^{n} \tilde{X}_{1,i} + \hat{\gamma}_2 \sum_{i=1}^{n} \tilde{X}_{1,i} X_{2,i} + \ldots + \hat{\gamma}_k \sum_{i=1}^{n} \tilde{X}_{1,i} X_{k,i} + \sum_{i=1}^{n} \tilde{X}_{1,i} \tilde{X}_{1,i} \\
&= \hat{\gamma}_0 \cdot 0 + \hat{\gamma}_2 \cdot 0 + \ldots + \hat{\gamma}_k \cdot 0 + \sum_{i=1}^{n} \tilde{X}_{1,i}^2 = \sum_{i=1}^{n} \tilde{X}_{1,i}^2,
\end{aligned}
$$

because of the normal equations for the X_1 regression.

Proof of the partitioned regression result (step 4)

Lastly, because $\hat{U}$ are the fitted residuals from the regression of Y against all X's:

$$\sum_{i=1}^{n} \hat{U}_i = \sum_{i=1}^{n} \hat{U}_i X_{1,i} = \ldots = \sum_{i=1}^{n} \hat{U}_i X_{k,i} = 0.$$

$$\begin{aligned}\sum_{i=1}^{n} \tilde{X}_{1,i}\hat{U}_i &= \sum_{i=1}^{n} \left(X_{1,i} - \hat{\gamma}_0 - \hat{\gamma}_2 X_{2,i} - \ldots - \hat{\gamma}_k X_{k,i}\right) \hat{U}_i \\ &= \sum_{i=1}^{n} X_{1,i}\hat{U}_i - \hat{\gamma}_0 \sum_{i=1}^{n} \hat{U}_i - \hat{\gamma}_2 \sum_{i=1}^{n} X_{2,i}\hat{U}_i - \ldots - \hat{\gamma}_k \sum_{i=1}^{n} X_{k,i}\hat{U}_i \\ &= 0 - \hat{\gamma}_0 \cdot 0 - \hat{\gamma}_2 \cdot 0 - \ldots - \hat{\gamma}_k \cdot 0 = 0.\end{aligned}$$

"Partialling out"

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n \tilde{X}_{1,i} Y_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2}$$

1. First, we regress X_1 against the rest of the regressors (and a constant) and keep $\tilde{X}_1$ which is the "part" of X_1 that is uncorrelated with other regressors (in sample, or orthogonal to other regressors).
2. Then, to obtain $\hat{\beta}_1$, we regress Y against $\tilde{X}_1$ which is "clean" from correlation with other regressors (no intercept).
3. $\hat{\beta}_1$ measures the effect of X_1 after the effects of $X_2, \ldots, X_k$ have been partialled out or netted out.