Lecture 9: Multiple regression

Economics 326 — Introduction to Econometrics II

Vadim Marmer, UBC

April 5, 2026

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 on** X_1:

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

- Since Y depends on X_2 as well,

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

- Assume that $\mathrm{E}\left[U_i \mid \mathbf{X}\right] = 0$, where $\mathbf{X} = \{(X_{1,i}, X_{2,i}) : i = 1, \ldots, n\}$. Then:

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

- This bias is called **omitted variable bias** because it arises from omitting X_2 from the regression.
- The exception (no omitted variable bias) is when X_1 and X_2 are "orthogonal":

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

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 Y 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: $\left\{\left(Y_i, X_{1,i}, X_{2,i}, \dots, X_{k,i}\right) : i = 1, \dots, n\right\}$.
- The model:

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

- In the general model, $\mathbf{X}$ denotes the full collection of regressors $\{X_{j,i} : j = 1, \dots, k;\ i = 1, \dots, n\}$.
- We also assume **no multicollinearity**: None of the regressors is constant, and there are no exact **linear** relationships among the regressors.

Interpretation of the coefficients

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

- 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 $X_2, \dots, X_k$).

$$\Delta Y = \beta_0 + \beta_1 \Delta X_1 + \beta_2 X_2 + \dots + \beta_k X_k + U.$$

- 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.
- Chandra et al., *Pediatrics*, 2008. "Does Watching Sex on Television Predict Teen Pregnancy?" National longitudinal survey of 718 youths (aged 12–17 at baseline).

$$\begin{aligned}\text{Teen Pregnancy} &= \beta_0 + \beta_1 \times \text{Exposure to Sex on TV} \\ &+ \beta_2 \times \text{Total TV} + U.\end{aligned}$$

- If we want to see the effect of Exposure, we have to increase the Total TV variable by the same amount as well.
- Otherwise, it is an effect of increasing sexual content and **decreasing** non-sexual content at the same time.
- Their estimates: $\hat{\beta}_1$ and $\hat{\beta}_2$ are of similar magnitude and **opposite** signs ($\hat{\beta}_1 = 0.44$ and $\hat{\beta}_2 = -0.42$).
- Alternative explanation: TV with no sexual content (cartoons, etc.) is **negatively** associated with teen pregnancy.

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**: $dY_i/dX_i = \beta_1$, which is **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:

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

  where 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, the return to education depends on the 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 sum of squared errors:

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

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

- The partial derivative with respect to b_0 is

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

- 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 the $\hat{U}$'s and the regressors are **orthogonal** (uncorrelated in the sample).

Partitioned regression

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

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

  - Here, $\hat{\gamma}_0, \hat{\gamma}_2, \ldots, \hat{\gamma}_k$ 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

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

- This formula requires $\sum_{i=1}^n \tilde{X}_{1,i}^2 \neq 0$. Under perfect multicollinearity, X_1 is a linear combination of the other regressors, so all residuals $\tilde{X}_{1,i} = 0$ and the OLS estimator does not exist.

Proof of the partitioned regression result

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

$$\begin{aligned}
&\frac{\sum_{i=1}^n \tilde{X}_{1,i} Y_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2} \\
&= \frac{\sum_{i=1}^n \tilde{X}_{1,i} \left(\hat{\beta}_0 + \hat{\beta}_1 X_{1,i} + \hat{\beta}_2 X_{2,i} + ... + \hat{\beta}_k X_{k,i} + \hat{U}_i\right)}{\sum_{i=1}^n \tilde{X}_{1,i}^2} \\
&= \hat{\beta}_0 \frac{\sum_{i=1}^n \tilde{X}_{1,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} + \hat{\beta}_1 \frac{\sum_{i=1}^n \tilde{X}_{1,i} X_{1,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} \\
&\quad + \hat{\beta}_2 \frac{\sum_{i=1}^n \tilde{X}_{1,i} X_{2,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} + ... + \hat{\beta}_k \frac{\sum_{i=1}^n \tilde{X}_{1,i} X_{k,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} + \frac{\sum_{i=1}^n \tilde{X}_{1,i} \hat{U}_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2}
\end{aligned}$$

- We will show that:
  1. $\sum_{i=1}^n \tilde{X}_{1,i} = 0$.
  2. $\sum_{i=1}^n \tilde{X}_{1,i} X_{2,i} = ... = \sum_{i=1}^n \tilde{X}_{1,i} X_{k,i} = 0$.
  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:

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

  where $\hat{\gamma}_0, \hat{\gamma}_2, ..., \hat{\gamma}_k$ 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 the $\tilde{X}_{1,i}$ are the fitted OLS residuals from the regression of X_1 on $X_2, \dots, X_k$:

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

(This follows from the **normal equations** for the X_1 regression.)

Proof of the partitioned regression result (step 4)

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

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

- Therefore,

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

"Partialling out"

- Recall:

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

1. First, we regress X_1 on the remaining regressors (and a constant) and keep $\tilde{X}_1$, which is the "part" of X_1 that is **uncorrelated** with the other regressors (in the sample, or orthogonal to the other regressors).
2. Then, to obtain $\hat{\beta}_1$, we regress Y on $\tilde{X}_1$, which is "clean" of correlation with the other regressors (no intercept).

- $\hat{\beta}_1$ measures the effect of X_1 after the effects of $X_2, \ldots, X_k$ have been **partialled out** or **netted out**.