Lecture 11: Properties of OLS in multiple regression

Economics 326 — Introduction to Econometrics II

Vadim Marmer, UBC

April 5, 2026

Multiple regression and OLS

- Consider the multiple regression model with k regressors:

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

- Let $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ be the OLS estimators: if

$$\hat{U}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - \hat{\beta}_2 X_{2,i} - \dots - \hat{\beta}_k X_{k,i},$$

then

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

Multiple regression and OLS

- As in Lecture 9, we can write $\hat{\beta}_1$ as

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

  - $\tilde{X}_{1,i}$ are the fitted OLS residuals: $\tilde{X}_{1,i} = X_{1,i} - \hat{\gamma}_0 - \hat{\gamma}_2 X_{2,i} - \dots - \hat{\gamma}_k X_{k,i}$.
  - $\hat{\gamma}_0, \hat{\gamma}_2, \dots, \hat{\gamma}_k$ are the OLS coefficients: $\sum_{i=1}^n \tilde{X}_{1,i} = \sum_{i=1}^n \tilde{X}_{1,i} X_{2,i} = \dots = \sum_{i=1}^n \tilde{X}_{1,i} X_{k,i} = 0$.

- Similarly, we can write $\hat{\beta}_2$ as

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n \tilde{X}_{2,i} Y_i}{\sum_{i=1}^n \tilde{X}_{2,i}^2}, \text{ where}$$

  - $\tilde{X}_{2,i}$ are the fitted OLS residuals: $\tilde{X}_{2,i} = X_{2,i} - \hat{\delta}_0 - \hat{\delta}_1 X_{1,i} - \hat{\delta}_3 X_{3,i} - \dots - \hat{\delta}_k X_{k,i}$.
  - $\hat{\delta}_0, \hat{\delta}_1, \hat{\delta}_3, \dots, \hat{\delta}_k$ are the OLS coefficients: $\sum_{i=1}^n \tilde{X}_{2,i} = \sum_{i=1}^n \tilde{X}_{2,i} X_{1,i} = \sum_{i=1}^n \tilde{X}_{2,i} X_{3,i} = \dots = \sum_{i=1}^n \tilde{X}_{2,i} X_{k,i} = 0$.

The OLS estimators are linear

- Consider $\hat{\beta}_1$:

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

  where

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

- Recall that $\tilde{X}_1$ are the residuals from a regression of X_1 on $X_2, \ldots, X_k$ and a constant, and therefore $w_{1,i}$ depends only on $\mathbf{X}$.

Unbiasedness

- Suppose that
  1. $Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_k X_{k,i} + U_i$.
  2. **Conditional on $\mathbf{X}$**, $\mathrm{E}[U_i \mid \mathbf{X}] = 0$ for all i's.
  - Conditioning on $\mathbf{X}$ means that we condition on all regressors for all observations: $\mathbf{X} = \{(X_{1,i}, X_{2,i}, \ldots, X_{k,i}) : i = 1, \ldots, n\}$.
- Under the above assumptions, conditional on $\mathbf{X}$:

$$\begin{aligned}
\mathrm{E}\left[\hat{\beta}_0 \mid \mathbf{X}\right] &= \beta_0, \\
\mathrm{E}\left[\hat{\beta}_1 \mid \mathbf{X}\right] &= \beta_1, \\
&\vdots \\
\mathrm{E}\left[\hat{\beta}_k \mid \mathbf{X}\right] &= \beta_k.
\end{aligned}$$

Proof of unbiasedness

- Substituting Y_i and expanding:

$$\begin{aligned}
\hat{\beta}_1 = \frac{\sum_{i=1}^n \tilde{X}_{1,i} Y_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2} &= \frac{\sum_{i=1}^n \tilde{X}_{1,i} \left(\beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_k X_{k,i} + U_i\right)}{\sum_{i=1}^n \tilde{X}_{1,i}^2} \\
&= \beta_0 \frac{\sum_{i=1}^n \tilde{X}_{1,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} + \beta_1 \frac{\sum_{i=1}^n \tilde{X}_{1,i} X_{1,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} + \beta_2 \frac{\sum_{i=1}^n \tilde{X}_{1,i} X_{2,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} \\
&\quad + \ldots + \beta_k \frac{\sum_{i=1}^n \tilde{X}_{1,i} X_{k,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} + \frac{\sum_{i=1}^n \tilde{X}_{1,i} U_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2}.
\end{aligned}$$

- Using the partitioned regression results from Lecture 9:

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

- Therefore,

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n} \tilde{X}_{1,i} U_i}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2}.$$

Proof of unbiasedness

- We have

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n} \tilde{X}_{1,i} U_i}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2}.$$

- **Conditional on X**,

$$\mathrm{E}\left[U_i \mid \mathbf{X}\right] = 0.$$

- Therefore, **conditional on X**,

$$\begin{aligned}
\mathrm{E}\left[\hat{\beta}_1 \mid \mathbf{X}\right] &= \mathrm{E}\left[\beta_1 + \frac{\sum_{i=1}^{n} \tilde{X}_{1,i} U_i}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2} \mid \mathbf{X}\right] \\
&= \beta_1 + \frac{\sum_{i=1}^{n} \tilde{X}_{1,i} \mathrm{E}\left[U_i \mid \mathbf{X}\right]}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2} \\
&= \beta_1.
\end{aligned}$$

Conditional variance of the OLS estimators

- Suppose that:
  1. $Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_k X_{k,i} + U_i$.
  2. **Conditional on X**, $\mathrm{E}\left[U_i \mid \mathbf{X}\right] = 0$ for all i's.
  3. **Conditional on X**, $\mathrm{E}\left[U_i^2 \mid \mathbf{X}\right] = \sigma^2$ for all i's.
  4. **Conditional on X**, $\mathrm{E}\left[U_i U_j \mid \mathbf{X}\right] = 0$ for all $i \neq j$.
- Denote by $SSR_1 = \sum_{i=1}^{n} \tilde{X}_{1,i}^2$ the residual sum-of-squares from regressing X_1 on a constant and the other regressors. The conditional variance of $\hat{\beta}_1$ given $\mathbf{X}$ is

$$\mathrm{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right) = \frac{\sigma^2}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2} = \frac{\sigma^2}{SSR_1}.$$

- **Gauss-Markov Theorem:** Under Assumptions 1–4, the OLS estimators are **BLUE**.

Derivation of the conditional variance

- We have $\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n} \tilde{X}_{1,i} U_i}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2}$.

- Conditional on $\mathbf{X}$,

$$
\begin{aligned}
\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right) &= \mathrm{E}\left[\left(\hat{\beta}_1 - \mathrm{E}\left[\hat{\beta}_1 \mid \mathbf{X}\right]\right)^2 \mid \mathbf{X}\right] \\
&= \mathrm{E}\left[\left(\frac{\sum_{i=1}^n \tilde{X}_{1,i} U_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2}\right)^2 \mid \mathbf{X}\right] \\
&= \frac{1}{\left(\sum_{i=1}^n \tilde{X}_{1,i}^2\right)^2} \mathrm{E}\left[\left(\sum_{i=1}^n \tilde{X}_{1,i} U_i\right)^2 \mid \mathbf{X}\right] \\
&= \frac{1}{\left(\sum_{i=1}^n \tilde{X}_{1,i}^2\right)^2} \left(\sum_{i=1}^n \tilde{X}_{1,i}^2 \sigma^2 + \sum_{i \neq j} \tilde{X}_{1,i} \tilde{X}_{1,j} \cdot 0\right) \\
&= \frac{\sigma^2 \sum_{i=1}^n \tilde{X}_{1,i}^2}{\left(\sum_{i=1}^n \tilde{X}_{1,i}^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^n \tilde{X}_{1,i}^2} = \frac{\sigma^2}{SSR_1}.
\end{aligned}
$$

Conditional covariance of the OLS estimators

- Consider $\hat{\beta}_1$ and $\hat{\beta}_2$:

$$
\begin{aligned}
\hat{\beta}_1 &= \beta_1 + \frac{\sum_{i=1}^n \tilde{X}_{1,i} U_i}{SSR_1}, \\
\hat{\beta}_2 &= \beta_2 + \frac{\sum_{i=1}^n \tilde{X}_{2,i} U_i}{SSR_2},
\end{aligned}
$$

  where $SSR_2 = \sum_{i=1}^n \tilde{X}_{2,i}^2$ is the residual sum-of-squares from regressing X_2 on a constant and $X_1, X_3, \ldots, X_k$.
- We will show that given Assumptions 1–4, **conditional on $\mathbf{X}$**:

$$
\operatorname{Cov}\left(\hat{\beta}_1, \hat{\beta}_2 \mid \mathbf{X}\right) = \sigma^2 \frac{\sum_{i=1}^n \tilde{X}_{1,i} \tilde{X}_{2,i}}{SSR_1 \cdot SSR_2}
$$

Conditional covariance of the OLS estimators

Conditional on $\mathbf{X}$,

$$
\begin{aligned}
&\operatorname{Cov}\left(\hat{\beta}_1, \hat{\beta}_2 \mid \mathbf{X}\right) \\
&= \mathrm{E}\left[\left(\hat{\beta}_1 - \mathrm{E}\left[\hat{\beta}_1 \mid \mathbf{X}\right]\right)\left(\hat{\beta}_2 - \mathrm{E}\left[\hat{\beta}_2 \mid \mathbf{X}\right]\right) \mid \mathbf{X}\right] \\
&= \frac{1}{SSR_1 \cdot SSR_2} \mathrm{E}\left[\left(\sum_{i=1}^n \tilde{X}_{1,i} U_i\right)\left(\sum_{i=1}^n \tilde{X}_{2,i} U_i\right) \mid \mathbf{X}\right] \\
&= \frac{1}{SSR_1 \cdot SSR_2} \left(\sum_{i=1}^n \tilde{X}_{1,i} \tilde{X}_{2,i} \sigma^2 + \sum_{i \neq j} \tilde{X}_{1,i} \tilde{X}_{2,j} \cdot 0\right) \\
&= \sigma^2 \frac{\sum_{i=1}^n \tilde{X}_{1,i} \tilde{X}_{2,i}}{SSR_1 \cdot SSR_2}.
\end{aligned}
$$

Normality of the OLS estimators

- In addition to Assumptions 1–4, assume that **conditional on $\mathbf{X}$**, U_i's are jointly normally distributed.

- $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ are linear estimators:

$$\hat{\beta}_j = \sum_{i=1}^{n} w_{j,i} Y_i = \beta_j + \sum_{i=1}^{n} w_{j,i} U_i,$$

where

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

and $\tilde{X}_{j,i}$ are the residuals from the regression of X_j on the rest of the regressors.

- It follows that $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ are jointly normally distributed (conditional on $\mathbf{X}$).

Inclusion of irrelevant regressors: No bias

- Suppose that the true model is $Y_i = \beta_0 + \beta_1 X_{1,i} + U_i$.
- We could estimate β_1 by

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

- Suppose that instead we regress Y on a constant, X_1, and $k-1$ additional regressors $X_2, \ldots, X_k$, i.e., we estimate β_1 by

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

- We have

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

- Since conditional on $\mathbf{X}$, $\mathrm{E}\left[U_i \mid \mathbf{X}\right] = 0$, $\tilde{\beta}_1$ is **unbiased**!

Inclusion of irrelevant regressors: Variance inflation

- When $Y_i = \beta_0 + \beta_1 X_{1,i} + U_i$,

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} \left(X_{1,i} - \bar{X}_1\right) Y_i}{\sum_{i=1}^{n} \left(X_{1,i} - \bar{X}_1\right)^2} \quad \text{and} \\ \tilde{\beta}_1 &= \frac{\sum_{i=1}^{n} \tilde{X}_{1,i} Y_i}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2}\end{aligned}$$

are both unbiased.

- Their variances, conditional on $\mathbf{X}$:

$$\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right) = \frac{\sigma^2}{\sum_{i=1}^n \left(X_{1,i} - \bar{X}_1\right)^2} \quad \text{and}$$
$$\operatorname{Var}\left(\tilde{\beta}_1 \mid \mathbf{X}\right) = \frac{\sigma^2}{SSR_1}.$$

- In the short regression, X_1 is regressed on a constant only, so $SSR_1 = \sum_{i=1}^n \left(X_{1,i} - \bar{X}_1\right)^2$.
- In the long regression, $X_2, \ldots, X_k$ are added. From Lecture 10, this cannot increase SSR_1, so

$$\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right) \leq \operatorname{Var}\left(\tilde{\beta}_1 \mid \mathbf{X}\right).$$

- Including irrelevant regressors inflates the variance of $\hat{\beta}_1$.

Variance and the number of regressors

- Recall the variance formula:

$$\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right) = \frac{\sigma^2}{SSR_1}.$$

- When we add a new regressor to the model, two quantities may change:
    1. σ^2 (the error variance): if the new regressor genuinely explains variation in Y, including it in the model moves that variation from U_i into the explained part, reducing $\operatorname{Var}(U_i \mid \mathbf{X}) = \sigma^2$
    2. SSR_1 (the variation in X_1 net of the other regressors): can only decrease or stay the same; stays the same only when the new regressor is uncorrelated with X_1

Case A: Irrelevant regressor

- Suppose the new regressor does not affect Y (its population coefficient is zero).
- σ^2 is unchanged, because the error variance is determined by the true data-generating process.
- SSR_1 decreases if the new regressor is correlated with X_1.
- Net effect: $\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right)$ **increases**.
- **Conclusion:** do not include irrelevant regressors.

Case B: Relevant regressor, uncorrelated with X_1

- Suppose the new regressor affects Y but is uncorrelated with X_1 after controlling for the other regressors.
- SSR_1 is approximately unchanged: the new regressor has no additional predictive power for X_1 beyond the existing regressors, so adding it to the auxiliary regression has a negligible effect on SSR_1 in large samples.
- σ^2 decreases, because the new regressor explains part of the variation in Y.
- Net effect: $\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right)$ **decreases**.
- **Conclusion:** always include such regressors.

Case C: Relevant regressor, correlated with X_1

- Two opposing forces:
  1. σ^2 decreases (the new regressor explains variation in Y)
  2. SSR_1 decreases (the new regressor is correlated with X_1)
- Net effect on $\mathrm{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right)$ is **ambiguous**: both the numerator and the denominator of σ^2/SSR_1 shrink, so the ratio could go either way.
- However, omitting a relevant regressor introduces **omitted variable bias** (see Lecture 9).
- **Conclusion:** always include relevant regressors, even if the variance of $\hat{\beta}_1$ may increase. An unbiased estimator with larger variance is preferable to a biased one.

Estimation of variances and covariances

- In $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1,i} + \hat{\beta}_2 X_{2,i} + \ldots + \hat{\beta}_k X_{k,i} + \hat{U}_i$,

$$
\begin{aligned}
\mathrm{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right) &= \frac{\sigma^2}{SSR_1}, \\
\mathrm{Cov}\left(\hat{\beta}_1, \hat{\beta}_2 \mid \mathbf{X}\right) &= \sigma^2 \frac{\sum_{i=1}^n \tilde{X}_{1,i}\tilde{X}_{2,i}}{SSR_1 \cdot SSR_2}.
\end{aligned}
$$

- Variances and covariances can be estimated by replacing σ^2 with

$$
s^2 = \frac{1}{n-k-1} \sum_{i=1}^n \hat{U}_i^2.
$$

- Estimated variance and covariance:

$$
\begin{aligned}
\widehat{\mathrm{Var}}\left(\hat{\beta}_1\right) &= \frac{s^2}{SSR_1}, \\
\widehat{\mathrm{Cov}}\left(\hat{\beta}_1, \hat{\beta}_2\right) &= s^2 \frac{\sum_{i=1}^n \tilde{X}_{1,i}\tilde{X}_{2,i}}{SSR_1 \cdot SSR_2}.
\end{aligned}
$$

Standard errors in terms of R-squared

- **Auxiliary R^2:** Let R_1^2 be the R-squared from regressing X_1 on a constant and $X_2, \ldots, X_k$. By definition of R-squared,

$$
SSR_1 = SST_1(1 - R_1^2), \quad \text{where } SST_1 = \sum_{i=1}^n (X_{1,i} - \bar{X}_1)^2.
$$

- **Adjusted R^2:** From $s^2 = s_Y^2(1 - \bar{R}^2)$, where $s_Y^2 = SST/(n-1)$ is the sample variance of Y,

$$
\mathrm{se}\left(\hat{\beta}_1\right) = \sqrt{\frac{s^2}{SSR_1}} = \sqrt{\frac{s_Y^2(1 - \bar{R}^2)}{SST_1(1 - R_1^2)}}.
$$

- Define $s_{X_1} = \sqrt{SST_1/(n-1)}$. Then

$$
\mathrm{se}\left(\hat{\beta}_1\right) = \frac{s_Y}{s_{X_1}} \cdot \frac{1}{\sqrt{n-1}} \cdot \sqrt{\frac{1 - \bar{R}^2}{1 - R_1^2}}.
$$

Three factors behind standard errors

- The SE formula has three interpretable factors:

$$\operatorname{se}\left(\hat{\beta}_1\right) = \underbrace{\frac{s_Y}{s_{X_1}}}_{\text{scaling}} \cdot \underbrace{\frac{1}{\sqrt{n-1}}}_{\text{sample size}} \cdot \underbrace{\sqrt{\frac{1-\bar{R}^2}{1-R_1^2}}}_{\text{fit vs. collinearity}}$$

1. s_Y/s_{X_1} — scaling: the ratio of sample standard deviations of Y and X_1
2. $1/\sqrt{n-1}$ — sample size effect; more data reduces SE
3. $\sqrt{(1-\bar{R}^2)/(1-R_1^2)}$ — fit vs. multicollinearity trade-off:
   - $(1-\bar{R}^2)$: unexplained variation in Y (adjusted for degrees of freedom); higher $\bar{R}^2$ reduces SE
   - $(1-R_1^2)$: unique variation in X_1; more collinearity (higher R_1^2) inflates SE

Connection to Cases A, B, C

- The SE formula clarifies the three cases:
  - **Case A** (irrelevant regressor): $\bar{R}^2$ decreases, R_1^2 may increase $\Longrightarrow$ SE increases
  - **Case B** (relevant, uncorrelated with X_1): $\bar{R}^2$ increases, $R_1^2 \approx$ unchanged $\Longrightarrow$ SE decreases
  - **Case C** (relevant, correlated with X_1): $\bar{R}^2$ increases but R_1^2 also increases $\Longrightarrow$ ambiguous
- **Remark.** An equivalent expression uses the unadjusted R^2:

$$\operatorname{se}\left(\hat{\beta}_1\right) = \frac{s_Y}{s_{X_1}} \cdot \frac{1}{\sqrt{n-k-1}} \cdot \sqrt{\frac{1-R^2}{1-R_1^2}}.$$

  This follows from $s^2 = SST(1-R^2)/(n-k-1)$, which keeps R^2 and the degrees-of-freedom correction separate.