Linear Regression with Multiple Regressors

Dr. Patrick Toche

Textbook:

James H. Stock and Mark W. Watson, *Introduction to Econometrics*, 4th Edition, Pearson.

Other references:

Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach*, 7th Edition, Cengage Learning.
Joshua D. Angrist and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, 1st Edition, Princeton University Press.

The textbook comes with online resources and study guides. Other references will be given from time to time.

Contents

In this lesson you will learn ...

- omitted variable bias
- least squares estimators in multiple regression
- measures of fit in multiple regression
- assumptions for causal inference
- distribution of OLS estimators
- multicollinearity
- control variables and conditional mean independence

In this lesson you will learn ...

- omitted variable bias
- least squares estimators in multiple regression
- measures of fit in multiple regression
- assumptions for causal inference
- distribution of OLS estimators
- multicollinearity
- control variables and conditional mean independence

In this lesson you will learn …

- omitted variable bias
- least squares estimators in multiple regression
- measures of fit in multiple regression
- assumptions for causal inference
- distribution of OLS estimators
- multicollinearity
- control variables and conditional mean independence

In this lesson you will learn ...

- omitted variable bias
- least squares estimators in multiple regression
- measures of fit in multiple regression
- assumptions for causal inference
- distribution of OLS estimators
- multicollinearity
- control variables and conditional mean independence

In this lesson you will learn ...

- omitted variable bias
- least squares estimators in multiple regression
- measures of fit in multiple regression
- assumptions for causal inference
- distribution of OLS estimators
- multicollinearity
- control variables and conditional mean independence

In this lesson you will learn ...

- omitted variable bias
- least squares estimators in multiple regression
- measures of fit in multiple regression
- assumptions for causal inference
- distribution of OLS estimators
- multicollinearity
- control variables and conditional mean independence

In this lesson you will learn ...

- omitted variable bias
- least squares estimators in multiple regression
- measures of fit in multiple regression
- assumptions for causal inference
- distribution of OLS estimators
- multicollinearity
- control variables and conditional mean independence

Omitted Variable Bias

Omitted Variable Bias

- **Definition:** Omitted variable bias occurs if (1) the omitted variable is correlated with the included regressor and (2) the omitted variable is a determinant of the dependent variable.
- **Omitted variable bias violates the first OLS assumption for causal inference:** E$[u_i|X_i = 0$. The error term u_i catches all factors other than X_i that are determinants of Y_i. If an omitted variable is a determinant of Y_i, then it is captured by the error term u_i. And if the omitted variable is correlated with X_i, then the error term u_i must be correlated with X_i and E$[u_i|X_i] \neq 0$.
- **Example:** Does listening to musing increase your IQ? One study from 1993 showed that students who take optional music or arts courses in high school have higher English and math test scores than those who don't. However, the correlation between testing well and taking art or music could be explained by an omitted variable. For instance, students who do better academically may also be more likely to take optional music classes, schools with a music curriculum could be better schools. By omitting factors such as the student's innate ability or the overall quality of the school, studying music appears to have an effect on test scores, but randomized controlled experiments have shown that it has no such effect.

Omitted Variable Bias

- **Definition:** Omitted variable bias occurs if (1) the omitted variable is correlated with the included regressor and (2) the omitted variable is a determinant of the dependent variable.
- **Omitted variable bias violates the first OLS assumption for causal inference:** E$[u_i|X_i = 0$. The error term u_i catches all factors other than X_i that are determinants of Y_i. If an omitted variable is a determinant of Y_i, then it is captured by the error term u_i. And if the omitted variable is correlated with X_i, then the error term u_i must be correlated with X_i and E$[u_i|X_i] \neq 0$.
- **Example:** Does listening to musing increase your IQ? One study from 1993 showed that students who take optional music or arts courses in high school have higher English and math test scores than those who don't. However, the correlation between testing well and taking art or music could be explained by an omitted variable. For instance, students who do better academically may also be more likely to take optional music classes, schools with a music curriculum could be better schools. By omitting factors such as the student's innate ability or the overall quality of the school, studying music appears to have an effect on test scores, but randomized controlled experiments have shown that it has no such effect.

Omitted Variable Bias

- **Definition:** Omitted variable bias occurs if (1) the omitted variable is correlated with the included regressor and (2) the omitted variable is a determinant of the dependent variable.
- **Omitted variable bias violates the first OLS assumption for causal inference:** E$[u_i|X_i = 0$. The error term u_i catches all factors other than X_i that are determinants of Y_i. If an omitted variable is a determinant of Y_i, then it is captured by the error term u_i. And if the omitted variable is correlated with X_i, then the error term u_i must be correlated with X_i and E$[u_i|X_i] \neq 0$.
- **Example:** Does listening to musing increase your IQ? One study from 1993 showed that students who take optional music or arts courses in high school have higher English and math test scores than those who don't. However, the correlation between testing well and taking art or music could be explained by an omitted variable. For instance, students who do better academically may also be more likely to take optional music classes, schools with a music curriculum could be better schools. By omitting factors such as the student's innate ability or the overall quality of the school, studying music appears to have an effect on test scores, but randomized controlled experiments have shown that it has no such effect.

Omitted Variable Bias

A Formula for the Omitted Variable Bias:

$$\hat{\beta}_1 \xrightarrow{p} \beta_1 + \rho_{Xu} \cdot \sigma_u / \sigma_X$$

- The notation $\xrightarrow{p}$ stands for "convergence in probability." As the sample size n increases, $\hat{\beta}_1 - \beta_1$ approaches $\rho_{Xu} \cdot \sigma_u / \sigma_X$ with increasingly high probability.
- The omitted variable bias does not disappear as the sample size is increased.
- With omitted variable bias, $\hat{\beta}_1$ is biased and inconsistent.
- The size of the bias depends on the correlation between the regressor and the error term ρ_{Xu}. The larger $|\rho_{Xu}|$, the larger the bias.
- The direction of the bias depends on the sign of the correlation.

Omitted Variable Bias

A Formula for the Omitted Variable Bias:

$$\hat{\beta}_1 \xrightarrow{p} \beta_1 + \rho_{Xu} \cdot \sigma_u / \sigma_X$$

- The notation $\xrightarrow{p}$ stands for "convergence in probability." As the sample size n increases, $\hat{\beta}_1 - \beta_1$ approaches $\rho_{Xu} \cdot \sigma_u / \sigma_X$ with increasingly high probability.
- The omitted variable bias does not disappear as the sample size is increased.
- With omitted variable bias, $\hat{\beta}_1$ is biased and inconsistent.
- The size of the bias depends on the correlation between the regressor and the error term ρ_{Xu}. The larger $|\rho_{Xu}|$, the larger the bias.
- The direction of the bias depends on the sign of the correlation.

Omitted Variable Bias

A Formula for the Omitted Variable Bias:

$$\hat{\beta}_1 \xrightarrow{p} \beta_1 + \rho_{Xu} \cdot \sigma_u / \sigma_X$$

- The notation $\xrightarrow{p}$ stands for "convergence in probability." As the sample size n increases, $\hat{\beta}_1 - \beta_1$ approaches $\rho_{Xu} \cdot \sigma_u / \sigma_X$ with increasingly high probability.
- The omitted variable bias does not disappear as the sample size is increased.
- With omitted variable bias, $\hat{\beta}_1$ is biased and inconsistent.
- The size of the bias depends on the correlation between the regressor and the error term ρ_{Xu}. The larger $|\rho_{Xu}|$, the larger the bias.
- The direction of the bias depends on the sign of the correlation.

A Formula for the Omitted Variable Bias:

$$\hat{\beta}_1 \xrightarrow{p} \beta_1 + \rho_{Xu} \cdot \sigma_u / \sigma_X$$

- The notation $\xrightarrow{p}$ stands for "convergence in probability." As the sample size n increases, $\hat{\beta}_1 - \beta_1$ approaches $\rho_{Xu} \cdot \sigma_u / \sigma_X$ with increasingly high probability.
- The omitted variable bias does not disappear as the sample size is increased.
- With omitted variable bias, $\hat{\beta}_1$ is biased and inconsistent.
- The size of the bias depends on the correlation between the regressor and the error term ρ_{Xu}. The larger $|\rho_{Xu}|$, the larger the bias.
- The direction of the bias depends on the sign of the correlation.

Omitted Variable Bias

A Formula for the Omitted Variable Bias:

$$\hat{\beta}_1 \xrightarrow{p} \beta_1 + \rho_{Xu} \cdot \sigma_u / \sigma_X$$

- The notation $\xrightarrow{p}$ stands for "convergence in probability." As the sample size n increases, $\hat{\beta}_1 - \beta_1$ approaches $\rho_{Xu} \cdot \sigma_u / \sigma_X$ with increasingly high probability.
- The omitted variable bias does not disappear as the sample size is increased.
- With omitted variable bias, $\hat{\beta}_1$ is biased and inconsistent.
- The size of the bias depends on the correlation between the regressor and the error term ρ_{Xu}. The larger $|\rho_{Xu}|$, the larger the bias.
- The direction of the bias depends on the sign of the correlation.

Omitted Variable Bias: Grouping Data

	Student-Teacher Ratio < 20		Student-Teacher Ratio ≥ 20		Difference in Test Scores, Low vs. High Student-Teacher Ratio	
	Average Test Score	*n*	Average Test Score	*n*	Difference	*t*-statistic
All districts	657.4	238	650.0	182	7.4	4.04
Percentage of English learners						
$< 1.9\%$	664.5	76	665.4	27	-0.9	-0.30
1.9–8.8%	665.2	64	661.8	44	3.3	1.13
8.8–23.0%	654.9	54	649.7	50	5.2	1.72
$> 23.0\%$	636.7	44	634.8	61	1.9	0.68

Differences in Test Scores for California School Districts with Low and High Student-Teacher Ratios, by the Percentage of English Learners in the District

Omitted Variable Bias: Grouping Data

English learners in California school districts:

- English learners: Students who are not native speakers and have not yet mastered English.
- **Districts with more English learners tend to have a higher student-teacher ratio.** The correlation between the student-teacher ratio and the percentage of English learners is 0.19.
- Because the student-teacher ratio and the percentage of English learners are correlated, it is possible that the OLS coefficient in the regression of test scores on the student-teacher ratio reflects that influence!
- The two conditions for an omitted variable bias are satisfied: **(1) The percentage of English learners is correlated with the student-teacher ratio. (2) The percentage of English learners is a determinant of test scores.** (students who are still learning English will do worse on standardized tests than native English speakers)
- The OLS estimator in the regression of test scores on the student-teacher ratio will reflect the influence of the omitted variable, the percentage of English learners.

Omitted Variable Bias: Grouping Data

English learners in California school districts:

- English learners: Students who are not native speakers and have not yet mastered English.
- **Districts with more English learners tend to have a higher student-teacher ratio.** The correlation between the student-teacher ratio and the percentage of English learners is 0.19.
- Because the student-teacher ratio and the percentage of English learners are correlated, it is possible that the OLS coefficient in the regression of test scores on the student-teacher ratio reflects that influence!
- The two conditions for an omitted variable bias are satisfied: **(1) The percentage of English learners is correlated with the student-teacher ratio. (2) The percentage of English learners is a determinant of test scores.** (students who are still learning English will do worse on standardized tests than native English speakers)
- The OLS estimator in the regression of test scores on the student-teacher ratio will reflect the influence of the omitted variable, the percentage of English learners.

Omitted Variable Bias: Grouping Data

English learners in California school districts:

- English learners: Students who are not native speakers and have not yet mastered English.
- **Districts with more English learners tend to have a higher student-teacher ratio.** The correlation between the student-teacher ratio and the percentage of English learners is 0.19.
- Because the student-teacher ratio and the percentage of English learners are correlated, it is possible that the OLS coefficient in the regression of test scores on the student-teacher ratio reflects that influence!
- The two conditions for an omitted variable bias are satisfied: **(1) The percentage of English learners is correlated with the student-teacher ratio. (2) The percentage of English learners is a determinant of test scores.** (students who are still learning English will do worse on standardized tests than native English speakers)
- The OLS estimator in the regression of test scores on the student-teacher ratio will reflect the influence of the omitted variable, the percentage of English learners.

Omitted Variable Bias: Grouping Data

English learners in California school districts:

- English learners: Students who are not native speakers and have not yet mastered English.
- **Districts with more English learners tend to have a higher student-teacher ratio.** The correlation between the student-teacher ratio and the percentage of English learners is 0.19.
- Because the student-teacher ratio and the percentage of English learners are correlated, it is possible that the OLS coefficient in the regression of test scores on the student-teacher ratio reflects that influence!
- The two conditions for an omitted variable bias are satisfied: **(1) The percentage of English learners is correlated with the student-teacher ratio. (2) The percentage of English learners is a determinant of test scores.** (students who are still learning English will do worse on standardized tests than native English speakers)
- The OLS estimator in the regression of test scores on the student-teacher ratio will reflect the influence of the omitted variable, the percentage of English learners.

Omitted Variable Bias: Grouping Data

English learners in California school districts:

- English learners: Students who are not native speakers and have not yet mastered English.
- **Districts with more English learners tend to have a higher student-teacher ratio.** The correlation between the student-teacher ratio and the percentage of English learners is 0.19.
- Because the student-teacher ratio and the percentage of English learners are correlated, it is possible that the OLS coefficient in the regression of test scores on the student-teacher ratio reflects that influence!
- The two conditions for an omitted variable bias are satisfied: **(1) The percentage of English learners is correlated with the student-teacher ratio. (2) The percentage of English learners is a determinant of test scores.** (students who are still learning English will do worse on standardized tests than native English speakers)
- The OLS estimator in the regression of test scores on the student-teacher ratio will reflect the influence of the omitted variable, the percentage of English learners.

Omitted Variable Bias: Grouping Data

Addressing Omitted Variable Bias: Analysis by Quartile

- **Hold the "omitted variable" constant:** (1) Select a subset of districts that have the same fraction of English learners but have different class sizes. (2) For that subset of districts, the fraction of English learners is held constant. (3) Look at the effect within each quartile.
- **Result:** The overall effect of test scores is twice the effect of test scores within any quartile! This is because the districts with the most English learners tend to have both the highest student-teacher ratios and the lowest test scores.
- The districts with few English learners tend to have lower student-teacher ratios:
- **Once we hold the percentage of English learners constant, the difference in test scores between districts with high and low student-teacher ratios is half — or less than half — of the overall estimate of 7.4 points.**
- To see this, estimate the effect of class size on test scores by quartile.

Omitted Variable Bias: Grouping Data

Addressing Omitted Variable Bias: Analysis by Quartile

- **Hold the "omitted variable" constant:** (1) Select a subset of districts that have the same fraction of English learners but have different class sizes. (2) For that subset of districts, the fraction of English learners is held constant. (3) Look at the effect within each quartile.
- **Result:** The overall effect of test scores is twice the effect of test scores within any quartile! This is because the districts with the most English learners tend to have both the highest student-teacher ratios and the lowest test scores.
- The districts with few English learners tend to have lower student-teacher ratios:
  - 74% of the districts in the 1st quartile have small classes (76 districts of 103)
  - 42% of the districts in the 4th quartile have small classes (44 districts of 105)
- **Once we hold the percentage of English learners constant, the difference in test scores between districts with high and low student-teacher ratios is half — or less than half — of the overall estimate of 7.4 points.**
- To see this, estimate the effect of class size on test scores by quartile.

Omitted Variable Bias: Grouping Data

Addressing Omitted Variable Bias: Analysis by Quartile

- **Hold the "omitted variable" constant:** (1) Select a subset of districts that have the same fraction of English learners but have different class sizes. (2) For that subset of districts, the fraction of English learners is held constant. (3) Look at the effect within each quartile.
- **Result:** The overall effect of test scores is twice the effect of test scores within any quartile! This is because the districts with the most English learners tend to have both the highest student-teacher ratios and the lowest test scores.
- The districts with few English learners tend to have lower student-teacher ratios:
  - 74% of the districts in the 1st quartile have small classes (76 districts of 103)
  - 42% of the districts in the 4th quartile have small classes (44 districts of 105)
- **Once we hold the percentage of English learners constant, the difference in test scores between districts with high and low student-teacher ratios is half — or less than half — of the overall estimate of 7.4 points.**
- To see this, estimate the effect of class size on test scores by quartile.

Omitted Variable Bias: Grouping Data

Addressing Omitted Variable Bias: Analysis by Quartile

- **Hold the "omitted variable" constant:** (1) Select a subset of districts that have the same fraction of English learners but have different class sizes. (2) For that subset of districts, the fraction of English learners is held constant. (3) Look at the effect within each quartile.
- **Result:** The overall effect of test scores is twice the effect of test scores within any quartile! This is because the districts with the most English learners tend to have both the highest student-teacher ratios and the lowest test scores.
- The districts with few English learners tend to have lower student-teacher ratios:
  - 74% of the districts in the 1st quartile have small classes (76 districts of 103)
  - 42% of the districts in the 4th quartile have small classes (44 districts of 105)
- Once we hold the percentage of English learners constant, the difference in test scores between districts with high and low student-teacher ratios is half — or less than half — of the overall estimate of 7.4 points.
- To see this, estimate the effect of class size on test scores by quartile.

Omitted Variable Bias: Grouping Data

Addressing Omitted Variable Bias: Analysis by Quartile

- **Hold the "omitted variable" constant:** (1) Select a subset of districts that have the same fraction of English learners but have different class sizes. (2) For that subset of districts, the fraction of English learners is held constant. (3) Look at the effect within each quartile.
- **Result:** The overall effect of test scores is twice the effect of test scores within any quartile! This is because the districts with the most English learners tend to have both the highest student-teacher ratios and the lowest test scores.
- The districts with few English learners tend to have lower student-teacher ratios:
  - 74% of the districts in the 1st quartile have small classes (76 districts of 103)
  - 42% of the districts in the 4th quartile have small classes (44 districts of 105)
- Once we hold the percentage of English learners constant, the difference in test scores between districts with high and low student-teacher ratios is half — or less than half — of the overall estimate of 7.4 points.
- To see this, estimate the effect of class size on test scores by quartile.

Omitted Variable Bias: Grouping Data

Addressing Omitted Variable Bias: Analysis by Quartile

- **Hold the "omitted variable" constant:** (1) Select a subset of districts that have the same fraction of English learners but have different class sizes. (2) For that subset of districts, the fraction of English learners is held constant. (3) Look at the effect within each quartile.
- **Result:** The overall effect of test scores is twice the effect of test scores within any quartile! This is because the districts with the most English learners tend to have both the highest student-teacher ratios and the lowest test scores.
- The districts with few English learners tend to have lower student-teacher ratios:
  - 74% of the districts in the 1st quartile have small classes (76 districts of 103)
  - 42% of the districts in the 4th quartile have small classes (44 districts of 105)
- **Once we hold the percentage of English learners constant, the difference in test scores between districts with high and low student-teacher ratios is half — or less than half — of the overall estimate of 7.4 points.**
- To see this, estimate the effect of class size on test scores by quartile.

Omitted Variable Bias: Grouping Data

Addressing Omitted Variable Bias: Analysis by Quartile

- **Hold the "omitted variable" constant:** (1) Select a subset of districts that have the same fraction of English learners but have different class sizes. (2) For that subset of districts, the fraction of English learners is held constant. (3) Look at the effect within each quartile.
- **Result:** The overall effect of test scores is twice the effect of test scores within any quartile! This is because the districts with the most English learners tend to have both the highest student-teacher ratios and the lowest test scores.
- The districts with few English learners tend to have lower student-teacher ratios:
  - 74% of the districts in the 1st quartile have small classes (76 districts of 103)
  - 42% of the districts in the 4th quartile have small classes (44 districts of 105)
- **Once we hold the percentage of English learners constant, the difference in test scores between districts with high and low student-teacher ratios is half — or less than half — of the overall estimate of 7.4 points.**
- To see this, estimate the effect of class size on test scores by quartile.

Omitted Variable Bias: Grouping Data

	Student-Teacher Ratio < 20		Student-Teacher Ratio ≥ 20		Difference in Test Scores, Low vs. High Student-Teacher Ratio	
	Average Test Score	n	Average Test Score	n	Difference	t-statistic
All districts	657.4	238	650.0	182	7.4	4.04
Percentage of English learners						
$< 1.9\%$ more small classes than large classes in districts with few English learners	664.5	76	665.4	27	−0.9	−0.30
1.9–8.8%	665.2	64	661.8	44	3.3	1.13
8.8–23.0%	654.9	54	649.7	50	5.2	1.72
$> 23.0\%$ more large classes than small classes in districts with few English learners	636.7	44	634.8	61	1.9	0.68

Differences in Test Scores for California School Districts with Low and High Student-Teacher Ratios, by the Percentage of English Learners in the District

Omitted Variable Bias: Grouping Data

	Student–Teacher Ratio < 20		Student–Teacher Ratio ≥ 20		Difference in Test Scores, Low vs. High Student–Teacher Ratio	
	Average Test Score	*n*	Average Test Score	*n*	Difference	*t*-statistic
All districts — much larger effect on test scores when all districts are aggregated	657.4	238	650.0	182	$\boxed{7.4}$	4.04
Percentage of English learners						
$< 1.9\%$ — no effect on test scores for districts with few English learners	664.5	76	665.4	27	$\boxed{-0.9}$	−0.30
1.9–8.8%	665.2	64	661.8	44	3.3	1.13
8.8–23.0%	654.9	54	649.7	50	5.2	1.72
$> 23.0\%$ — somewhat smaller effect on test scores for districts with many English learners	636.7	44	634.8	61	$\boxed{1.9}$	0.68

Differences in Test Scores for California School Districts with Low and High Student-Teacher Ratios, by the Percentage of English Learners in the District

Omitted Variable Bias: Grouping Data

Addressing Omitted Variable Bias:

- To estimate the effect of the student-teacher ratio on test scores, we hold constant the percentage of English learners. Districts are divided into 8 groups: by quartile of the distribution of English learners across districts and by the student-teacher ratio (high vs low).
- Over the full sample of 420 districts, the average test score is 7.4 points higher in districts with a lower student-teacher ratio. The t-statistic is 4.04, so the null hypothesis that the mean test score is the same in the two groups is rejected at the 1% significance level.
- But different results emerge if the difference in test scores between districts with low and high ratios is broken down by the quartile of the percentage of English learners.
- Looking within quartiles of the percentage of English learners improves on the simple difference-of-means analysis. **But to estimate the effect on test scores of changing class size, holding constant the fraction of English learners, we must perform a multiple regression.**

Omitted Variable Bias: Grouping Data

Addressing Omitted Variable Bias:

- To estimate the effect of the student-teacher ratio on test scores, we hold constant the percentage of English learners. Districts are divided into 8 groups: by quartile of the distribution of English learners across districts and by the student-teacher ratio (high vs low).
- Over the full sample of 420 districts, the average test score is 7.4 points higher in districts with a lower student-teacher ratio. The t-statistic is 4.04, so the null hypothesis that the mean test score is the same in the two groups is rejected at the 1% significance level.
- But different results emerge if the difference in test scores between districts with low and high ratios is broken down by the quartile of the percentage of English learners.
  - 1st Quartile: The average test score was not appreciably different in districts with high vs low student-teacher ratio — the difference is small and in the opposite direction as the overall effect.
  - 2nd Quartile: The average test score was 3.3 points higher in districts with small class sizes.
  - 3rd Quartile: The average test score was 5.2 points higher.
  - 4th Quartile: The average test score was 1.9 points higher.
- Looking within quartiles of the percentage of English learners improves on the simple difference-of-means analysis. **But to estimate the effect on test scores of changing class size, holding constant the fraction of English learners, we must perform a multiple regression.**

Omitted Variable Bias: Grouping Data

Addressing Omitted Variable Bias:

- To estimate the effect of the student-teacher ratio on test scores, we hold constant the percentage of English learners. Districts are divided into 8 groups: by quartile of the distribution of English learners across districts and by the student-teacher ratio (high vs low).
- Over the full sample of 420 districts, the average test score is 7.4 points higher in districts with a lower student-teacher ratio. The t-statistic is 4.04, so the null hypothesis that the mean test score is the same in the two groups is rejected at the 1% significance level.
- But different results emerge if the difference in test scores between districts with low and high ratios is broken down by the quartile of the percentage of English learners.
  - 1st Quartile: The average test score was not appreciably different in districts with high vs low student-teacher ratio — the difference is small and in the opposite direction as the overall effect.
  - 2nd Quartile: The average test score was 3.3 points higher in districts with small class sizes.
  - 3rd Quartile: The average test score was 5.2 points higher.
  - 4th Quartile: The average test score was 1.9 points higher.
- Looking within quartiles of the percentage of English learners improves on the simple difference-of-means analysis. **But to estimate the effect on test scores of changing class size, holding constant the fraction of English learners, we must perform a multiple regression.**

Omitted Variable Bias: Grouping Data

Addressing Omitted Variable Bias:

- To estimate the effect of the student-teacher ratio on test scores, we hold constant the percentage of English learners. Districts are divided into 8 groups: by quartile of the distribution of English learners across districts and by the student-teacher ratio (high vs low).
- Over the full sample of 420 districts, the average test score is 7.4 points higher in districts with a lower student-teacher ratio. The t-statistic is 4.04, so the null hypothesis that the mean test score is the same in the two groups is rejected at the 1% significance level.
- But different results emerge if the difference in test scores between districts with low and high ratios is broken down by the quartile of the percentage of English learners.
  - 1st Quartile: The average test score was not appreciably different in districts with high vs low student-teacher ratio – the difference is small and in the opposite direction as the overall effect.
  - 2nd Quartile: The average test score was 3.3 points higher in districts with small class sizes.
  - 3rd Quartile: The average test score was 5.2 points higher.
  - 4th Quartile: The average test score was 1.9 points higher.
- Looking within quartiles of the percentage of English learners improves on the simple difference-of-means analysis. **But to estimate the effect on test scores of changing class size, holding constant the fraction of English learners, we must perform a multiple regression.**

Omitted Variable Bias: Grouping Data

Addressing Omitted Variable Bias:

- To estimate the effect of the student-teacher ratio on test scores, we hold constant the percentage of English learners. Districts are divided into 8 groups: by quartile of the distribution of English learners across districts and by the student-teacher ratio (high vs low).
- Over the full sample of 420 districts, the average test score is 7.4 points higher in districts with a lower student-teacher ratio. The t-statistic is 4.04, so the null hypothesis that the mean test score is the same in the two groups is rejected at the 1% significance level.
- But different results emerge if the difference in test scores between districts with low and high ratios is broken down by the quartile of the percentage of English learners.
  - 1st Quartile: The average test score was not appreciably different in districts with high vs low student-teacher ratio – the difference is small and in the opposite direction as the overall effect.
  - 2nd Quartile: The average test score was 3.3 points higher in districts with small class sizes.
  - 3rd Quartile: The average test score was 5.2 points higher.
  - 4th Quartile: The average test score was 1.9 points higher.
- Looking within quartiles of the percentage of English learners improves on the simple difference-of-means analysis. **But to estimate the effect on test scores of changing class size, holding constant the fraction of English learners, we must perform a multiple regression.**

Omitted Variable Bias: Grouping Data

Addressing Omitted Variable Bias:

- To estimate the effect of the student-teacher ratio on test scores, we hold constant the percentage of English learners. Districts are divided into 8 groups: by quartile of the distribution of English learners across districts and by the student-teacher ratio (high vs low).
- Over the full sample of 420 districts, the average test score is 7.4 points higher in districts with a lower student-teacher ratio. The t-statistic is 4.04, so the null hypothesis that the mean test score is the same in the two groups is rejected at the 1% significance level.
- But different results emerge if the difference in test scores between districts with low and high ratios is broken down by the quartile of the percentage of English learners.
  - 1st Quartile: The average test score was not appreciably different in districts with high vs low student-teacher ratio – the difference is small and in the opposite direction as the overall effect.
  - 2nd Quartile: The average test score was 3.3 points higher in districts with small class sizes.
  - 3rd Quartile: The average test score was 5.2 points higher.
  - 4th Quartile: The average test score was 1.9 points higher.
- Looking within quartiles of the percentage of English learners improves on the simple difference-of-means analysis. **But to estimate the effect on test scores of changing class size, holding constant the fraction of English learners, we must perform a multiple regression.**

Omitted Variable Bias: Grouping Data

Addressing Omitted Variable Bias:

- To estimate the effect of the student-teacher ratio on test scores, we hold constant the percentage of English learners. Districts are divided into 8 groups: by quartile of the distribution of English learners across districts and by the student-teacher ratio (high vs low).
- Over the full sample of 420 districts, the average test score is 7.4 points higher in districts with a lower student-teacher ratio. The t-statistic is 4.04, so the null hypothesis that the mean test score is the same in the two groups is rejected at the 1% significance level.
- But different results emerge if the difference in test scores between districts with low and high ratios is broken down by the quartile of the percentage of English learners.
  - 1st Quartile: The average test score was not appreciably different in districts with high vs low student-teacher ratio – the difference is small and in the opposite direction as the overall effect.
  - 2nd Quartile: The average test score was 3.3 points higher in districts with small class sizes.
  - 3rd Quartile: The average test score was 5.2 points higher.
  - 4th Quartile: The average test score was 1.9 points higher.
- Looking within quartiles of the percentage of English learners improves on the simple difference-of-means analysis. **But to estimate the effect on test scores of changing class size, holding constant the fraction of English learners, we must perform a multiple regression.**

Omitted Variable Bias: Grouping Data

Addressing Omitted Variable Bias:

- To estimate the effect of the student-teacher ratio on test scores, we hold constant the percentage of English learners. Districts are divided into 8 groups: by quartile of the distribution of English learners across districts and by the student-teacher ratio (high vs low).
- Over the full sample of 420 districts, the average test score is 7.4 points higher in districts with a lower student-teacher ratio. The t-statistic is 4.04, so the null hypothesis that the mean test score is the same in the two groups is rejected at the 1% significance level.
- But different results emerge if the difference in test scores between districts with low and high ratios is broken down by the quartile of the percentage of English learners.
  - 1st Quartile: The average test score was not appreciably different in districts with high vs low student-teacher ratio – the difference is small and in the opposite direction as the overall effect.
  - 2nd Quartile: The average test score was 3.3 points higher in districts with small class sizes.
  - 3rd Quartile: The average test score was 5.2 points higher.
  - 4th Quartile: The average test score was 1.9 points higher.
- Looking within quartiles of the percentage of English learners improves on the simple difference-of-means analysis. **But to estimate the effect on test scores of changing class size, holding constant the fraction of English learners, we must perform a multiple regression.**

Multiple Regression

Population Regression Line

Population Regression:

- The population regression function in the multiple regression model is a model of the conditional expectation of Y_i:

$$\mathsf{E}[Y_i | X_{1i} = x_1, X_{2i} = x_2] = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

- **Example:** Effect of class size in California districts: Multiple regression allows us to isolate the effect on test scores (Y_i) of the student-teacher ratio (X_{1i}), while holding other regressors constant, in particular the percentage of students in the district who are English learners (X_{2i}).

Population Regression Line

Population Regression:

- The population regression function in the multiple regression model is a model of the conditional expectation of Y_i:

$$\mathsf{E}[Y_i|X_{1i} = x_1, X_{2i} = x_2] = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

- **Example:** Effect of class size in California districts: Multiple regression allows us to isolate the effect on test scores (Y_i) of the student-teacher ratio (X_{1i}), while holding other regressors constant, in particular the percentage of students in the district who are English learners (X_{2i}).

Interpreting the slope coefficient:

- The slope coefficient β_1 is the predicted difference in Y between two observations with a unit difference in X_1, holding X_2 constant.
- Consider a change in X_{1i} holding X_{2i} constant:

$$Y_i + \Delta Y_i = \beta_0 + \beta_1 (X_{1i} + \Delta X_{1i}) + \beta_2 X_{2i}$$

$$Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} = u_i = \beta_1 \Delta X_{1i} - \Delta Y_i$$

- Calculate the expected value of Y_i evaluated at $X_{1i} = x_1$ and $X_{2i} = x_2$:

$$\begin{aligned} \mathsf{E}[Y_i | X_{1i} = x_1, X_{2i} = x_2] - \beta_0 - \beta_1 x_1 - \beta_2 x_2 &= \mathsf{E}[u_i | X_{1i} = x_1, X_{2i} = x_2] = 0 \\ &= \beta_1 \Delta X_1 - \Delta Y \Big|_{X_1 = x_1, X_2 = x_2} \\ \implies \beta_1 &= \frac{\Delta Y}{\Delta X_1}\Big|_{X_1 = x_1, X_2 = x_2} \end{aligned}$$

Population Regression Line

Interpreting the slope coefficient:

- The slope coefficient β_1 is the predicted difference in Y between two observations with a unit difference in X_1, holding X_2 constant.
- Consider a change in X_{1i} holding X_{2i} constant:

$$Y_i + \Delta Y_i = \beta_0 + \beta_1 (X_{1i} + \Delta X_{1i}) + \beta_2 X_{2i}$$
$$Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} = u_i = \beta_1 \Delta X_{1i} - \Delta Y_i$$

- Calculate the expected value of Y_i evaluated at $X_{1i} = x_1$ and $X_{2i} = x_2$:

$$\begin{aligned} \mathsf{E}[Y_i | X_{1i} = x_1, X_{2i} = x_2] - \beta_0 - \beta_1 x_1 - \beta_2 x_2 &= \mathsf{E}[u_i | X_{1i} = x_1, X_{2i} = x_2] = 0 \\ &= \beta_1 \Delta X_1 - \Delta Y \Big|_{X_1 = x_1, X_2 = x_2} \\ \Longrightarrow \beta_1 &= \frac{\Delta Y}{\Delta X_1}\Big|_{X_1 = x_1, X_2 = x_2} \end{aligned}$$

Population Regression Line

Interpreting the slope coefficient:

- The slope coefficient β_1 is the predicted difference in Y between two observations with a unit difference in X_1, holding X_2 constant.
- Consider a change in X_{1i} holding X_{2i} constant:

$$Y_i + \Delta Y_i = \beta_0 + \beta_1(X_{1i} + \Delta X_{1i}) + \beta_2 X_{2i}$$
$$Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} = u_i = \beta_1 \Delta X_{1i} - \Delta Y_i$$

- Calculate the expected value of Y_i evaluated at $X_{1i} = x_1$ and $X_{2i} = x_2$:

$$\begin{aligned} \mathsf{E}[Y_i | X_{1i} = x_1, X_{2i} = x_2] - \beta_0 - \beta_1 x_1 - \beta_2 x_2 = \mathsf{E}[u_i | X_{1i} = x_1, X_{2i} = x_2] &= 0 \\ &= \beta_1 \Delta X_1 - \Delta Y \Big|_{X_1 = x_1, X_2 = x_2} \\ \implies \beta_1 &= \frac{\Delta Y}{\Delta X_1}\Big|_{X_1 = x_1, X_2 = x_2} \end{aligned}$$

Least Squares Estimator

Least Squares Estimator

OLS Estimator:

- The method of Ordinary Least Squares (OLS) selects $k + 1$ estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$, to minimize the sum of the squared residuals from the regression function:

$$\min_{\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k} \sum_{i=1}^{n} \hat{u}_i^2$$

- The OLS estimates are computed from a sample of n observations for each of the k regressors $X_{1i}, \ldots, X_{ki}$ and the regressand Y_i.
- The n residuals $\hat{u}_i$ are:

$$\hat{u}_i = Y_i - \hat{Y}_i, i = 1, \ldots, n$$

- Equivalently:

$$\sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_{1i} - \beta_k X_{ki})^2 \to \min$$

Least Squares Estimator

OLS Estimator:

- The method of Ordinary Least Squares (OLS) selects $k+1$ estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$, to minimize the sum of the squared residuals from the regression function:

$$\min_{\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k} \sum_{i=1}^{n} \hat{u}_i^2$$

- The OLS estimates are computed from a sample of n observations for each of the k regressors $X_{1i}, \ldots, X_{ki}$ and the regressand Y_i.
- The n residuals $\hat{u}_i$ are:

$$\hat{u}_i = Y_i - \hat{Y}_i, i = 1, \ldots, n$$

- Equivalently:

$$\sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_{1i} - \beta_k X_{ki})^2 \to \min$$

Least Squares Estimator

OLS Estimator:

- The method of Ordinary Least Squares (OLS) selects $k+1$ estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$, to minimize the sum of the squared residuals from the regression function:

$$\min_{\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k} \sum_{i=1}^{n} \hat{u}_i^2$$

- The OLS estimates are computed from a sample of n observations for each of the k regressors $X_{1i}, \ldots, X_{ki}$ and the regressand Y_i.
- The n residuals $\hat{u}_i$ are:

$$\hat{u}_i = Y_i - \hat{Y}_i, i = 1, \ldots, n$$

- Equivalently:

$$\sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_{1i} - \beta_k X_{ki})^2 \to \min$$

Least Squares Estimator

OLS Estimator:

- The method of Ordinary Least Squares (OLS) selects $k+1$ estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$, to minimize the sum of the squared residuals from the regression function:

$$\min_{\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k} \sum_{i=1}^{n} \hat{u}_i^2$$

- The OLS estimates are computed from a sample of n observations for each of the k regressors $X_{1i}, \ldots, X_{ki}$ and the regressand Y_i.
- The n residuals $\hat{u}_i$ are:

$$\hat{u}_i = Y_i - \hat{Y}_i, i = 1, \ldots, n$$

- Equivalently:

$$\sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_{1i} - \beta_k X_{ki})^2 \rightarrow \min$$

Measures of Fit

Standard Error of the Regression

Standard Error of the Regression:

- A measure of the spread of the distribution of Y around the regression line. The standard error of the regression (SER) is the square-root of the mean squared residuals with an adjustment for degrees of freedom:

$$SSR = \sum_{i=1}^{n} \hat{u}_i^2$$

$$SER = \sqrt{\frac{SSR}{n-k-1}}$$

  where SSR stands for the sum of the squared residuals.
- $n-k-1$ adjusts for the downward bias introduced from estimating $k+1$ coefficients.
- A related measure is the root mean squared error (RMSE) — the square-root of the mean squared error (MSE), with no adjustment for degrees of freedom:

$$MSE = \frac{SSR}{n}$$

$$RMSE = \sqrt{MSE}$$

Standard Error of the Regression

Standard Error of the Regression:

- A measure of the spread of the distribution of Y around the regression line. The standard error of the regression (SER) is the square-root of the mean squared residuals with an adjustment for degrees of freedom:

$$SSR = \sum_{i=1}^{n} \hat{u}_i^2$$

$$SER = \sqrt{\frac{SSR}{n-k-1}}$$

where SSR stands for the sum of the squared residuals.

- $n-k-1$ adjusts for the downward bias introduced from estimating $k+1$ coefficients.
- A related measure is the root mean squared error (RMSE) — the square-root of the mean squared error (MSE), with no adjustment for degrees of freedom:

$$MSE = \frac{SSR}{n}$$

$$RMSE = \sqrt{MSE}$$

Standard Error of the Regression

Standard Error of the Regression:

- A measure of the spread of the distribution of Y around the regression line. The standard error of the regression (SER) is the square-root of the mean squared residuals with an adjustment for degrees of freedom:

$$SSR = \sum_{i=1}^{n} \hat{u}_i^2$$

$$SER = \sqrt{\frac{SSR}{n-k-1}}$$

where SSR stands for the sum of the squared residuals.

- $n-k-1$ adjusts for the downward bias introduced from estimating $k+1$ coefficients.
- A related measure is the root mean squared error (RMSE) — the square-root of the mean squared error (MSE), with no adjustment for degrees of freedom:

$$MSE = \frac{SSR}{n}$$

$$RMSE = \sqrt{MSE}$$

Coefficient of Determination R^2

Coefficient of Determination

- The regression R^2 is the fraction of the sample variance of Y_i explained by the regressors.
- Equivalently, R^2 is 1 minus the fraction of the variance of Y_i not explained by the regressors.

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

where

$$TSS = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2$$

$$ESS = \sum_{i=1}^{n} \left(\hat{Y}_i - \overline{Y} \right)^2$$

$$SSR = \sum_{i=1}^{n} \left(Y_i - \overline{Y} \right)^2$$

Coefficient of Determination R^2

Coefficient of Determination

- The regression R^2 is the fraction of the sample variance of Y_i explained by the regressors.
- Equivalently, R^2 is 1 minus the fraction of the variance of Y_i not explained by the regressors.

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

where

$$TSS = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$$

$$ESS = \sum_{i=1}^{n} \left(\hat{Y}_i - \overline{Y}\right)^2$$

$$SSR = \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2$$

Adjusted R^2

Adjusted R^2:

$$\bar{R}^2 = 1 - \frac{n-1}{n-1-k}\frac{SSR}{TSS}$$
$$= 1 - \frac{s_{\hat{u}}^2}{s_Y^2}$$

- Because the R^2 increases when a new variable is added, an increase in the R^2 does not mean that adding a variable actually improves the fit of the model – R^2 gives an inflated estimate of how well the regression fits the data. The adjusted R^2, denoted $\bar{R}^2$, corrects that.
- Adding a regressor has two opposite effects on the R^2:
  - the SSR falls.
  - $(n-1)/(n-1-k)$ rises.
- By construction, $\bar{R}^2 \leq R^2$.
- A negative $\bar{R}^2$ is theoretically possible.

Adjusted R^2:

$$\bar{R}^2 = 1 - \frac{n-1}{n-1-k} \frac{SSR}{TSS}$$
$$= 1 - \frac{s_{\hat{u}}^2}{s_Y^2}$$

- Because the R^2 increases when a new variable is added, an increase in the R^2 does not mean that adding a variable actually improves the fit of the model — R^2 gives an inflated estimate of how well the regression fits the data. The adjusted R^2, denoted $\bar{R}^2$, corrects that.
- Adding a regressor has two opposite effects on the R^2:
  - the SSR falls.
  - $(n-1)/(n-1-k)$ rises.
- By construction, $\bar{R}^2 \leq R^2$.
- A negative $\bar{R}^2$ is theoretically possible.

Adjusted R^2

Adjusted $\boldsymbol{R^2}$:

$$\bar{R}^2 = 1 - \frac{n-1}{n-1-k}\frac{SSR}{TSS}$$
$$= 1 - \frac{s_{\hat{u}}^2}{s_Y^2}$$

- Because the R^2 increases when a new variable is added, an increase in the R^2 does not mean that adding a variable actually improves the fit of the model — R^2 gives an inflated estimate of how well the regression fits the data. The adjusted R^2, denoted $\bar{R}^2$, corrects that.
- Adding a regressor has two opposite effects on the R^2:
  - the SSR falls.
  - $(n-1)/(n-1-k)$ rises.
- By construction, $\bar{R}^2 \leq R^2$.
- A negative $\bar{R}^2$ is theoretically possible.

Adjusted R^2

Adjusted R^2:

$$\bar{R}^2 = 1 - \frac{n-1}{n-1-k}\frac{SSR}{TSS}$$
$$= 1 - \frac{s_{\hat{u}}^2}{s_Y^2}$$

- Because the R^2 increases when a new variable is added, an increase in the R^2 does not mean that adding a variable actually improves the fit of the model — R^2 gives an inflated estimate of how well the regression fits the data. The adjusted R^2, denoted $\bar{R}^2$, corrects that.
- Adding a regressor has two opposite effects on the R^2:
  - the SSR falls.
  - $(n-1)/(n-1-k)$ rises.
- By construction, $\bar{R}^2 \leq R^2$.
- A negative $\bar{R}^2$ is theoretically possible.

Adjusted R^2

Adjusted R^2:

$$\bar{R}^2 = 1 - \frac{n-1}{n-1-k}\frac{SSR}{TSS}$$
$$= 1 - \frac{s_{\hat{u}}^2}{s_Y^2}$$

- Because the R^2 increases when a new variable is added, an increase in the R^2 does not mean that adding a variable actually improves the fit of the model — R^2 gives an inflated estimate of how well the regression fits the data. The adjusted R^2, denoted $\bar{R}^2$, corrects that.
- Adding a regressor has two opposite effects on the R^2:
  - the SSR falls.
  - $(n-1)/(n-1-k)$ rises.
- By construction, $\bar{R}^2 \leq R^2$.
- A negative $\bar{R}^2$ is theoretically possible.

Adjusted R^2

Adjusted R^2:

$$\bar{R}^2 = 1 - \frac{n-1}{n-1-k}\frac{SSR}{TSS}$$
$$= 1 - \frac{s_{\hat{u}}^2}{s_Y^2}$$

- Because the R^2 increases when a new variable is added, an increase in the R^2 does not mean that adding a variable actually improves the fit of the model – R^2 gives an inflated estimate of how well the regression fits the data. The adjusted R^2, denoted $\bar{R}^2$, corrects that.
- Adding a regressor has two opposite effects on the R^2:
  - the SSR falls.
  - $(n-1)/(n-1-k)$ rises.
- By construction, $\bar{R}^2 \leq R^2$.
- A negative $\bar{R}^2$ is theoretically possible.

Assumptions for Causal Inference

Least Squares Assumptions

Assumptions for causal inference:

- Causal inference is possible only under a strict set of assumptions.
  1. Errors have zero mean: $\mathrm{E}[\hat{u}_i | X_{1i} = x_1, \ldots, X_{ki} = x_k] = 0$. The conditional distribution of u_i given $X_{1i}, \ldots, X_{ki}$ has zero mean.
  2. Regressors are i.i.d.: $X_{1i}, \ldots, X_{ki} \sim$ i.i.d.. Random sampling implies the regressors are independently and identically distributed.
  3. Large outliers are unlikely: The regressand and regressors have finite kurtosis, $\mathrm{E}[Y^4] < \infty, \mathrm{E}[X_j^4] < \infty$.
  4. No perfect multicollinearity among the regressors.
- Under these assumptions, the least-squares estimators $\hat{\beta}_0, \ldots, \hat{\beta}_k$ are jointly normally distributed with

$$\hat{\beta}_j \sim \mathcal{N}(\beta_j, \sigma^2_{\beta_j}), j = 0, \ldots, k.$$

Least Squares Assumptions

Assumptions for causal inference:

- Causal inference is possible only under a strict set of assumptions.
  1. Errors have zero mean: $\mathsf{E}[\hat{u}_i | X_{1i} = x_1, \ldots, X_{ki} = x_k] = 0$. The conditional distribution of u_i given $X_{1i}, \ldots, X_{ki}$ has zero mean.
  2. Regressors are i.i.d.: $X_{1i}, \ldots, X_{ki} \sim$ i.i.d.. Random sampling implies the regressors are independently and identically distributed.
  3. Large outliers are unlikely: The regressand and regressors have finite kurtosis, $\mathsf{E}[Y^4] < \infty$, $\mathsf{E}[X_j^4] < \infty$.
  4. No perfect multicollinearity among the regressors.
- Under these assumptions, the least-squares estimators $\hat{\beta}_0, \ldots, \hat{\beta}_k$ are jointly normally distributed with

$$\hat{\beta}_j \sim \mathcal{N}(\beta_j, \sigma^2_{\beta_j}), j = 0, \ldots, k.$$

Least Squares Assumptions

Assumptions for causal inference:

- Causal inference is possible only under a strict set of assumptions.
  1. Errors have zero mean: $\mathsf{E}[\hat{u}_i | X_{1i} = x_1, \ldots, X_{ki} = x_k] = 0$. The conditional distribution of u_i given $X_{1i}, \ldots, X_{ki}$ has zero mean.
  2. Regressors are i.i.d.: $X_{1i}, \ldots, X_{ki} \sim$ i.i.d.. Random sampling implies the regressors are independently and identically distributed.
  3. Large outliers are unlikely: The regressand and regressors have finite kurtosis, $\mathsf{E}[Y^4] < \infty$, $\mathsf{E}[X_j^4] < \infty$.
  4. No perfect multicollinearity among the regressors.
- Under these assumptions, the least-squares estimators $\hat{\beta}_0, \ldots, \hat{\beta}_k$ are jointly normally distributed with

$$\hat{\beta}_j \sim \mathcal{N}(\beta_j, \sigma^2_{\hat{\beta}_j}), j = 0, \ldots, k.$$

Least Squares Assumptions

Assumptions for causal inference:

- Causal inference is possible only under a strict set of assumptions.
  1. Errors have zero mean: $\mathrm{E}[\hat{u}_i | X_{1i} = x_1, \ldots, X_{ki} = x_k] = 0$. The conditional distribution of u_i given $X_{1i}, \ldots, X_{ki}$ has zero mean.
  2. Regressors are i.i.d.: $X_{1i}, \ldots, X_{ki} \sim$ i.i.d.. Random sampling implies the regressors are independently and identically distributed.
  3. Large outliers are unlikely: The regressand and regressors have finite kurtosis, $\mathrm{E}[Y^4] < \infty$, $\mathrm{E}[X_j^4] < \infty$.
  4. No perfect multicollinearity among the regressors.
- Under these assumptions, the least-squares estimators $\hat{\beta}_0, \ldots, \hat{\beta}_k$ are jointly normally distributed with

$$\hat{\beta}_j \sim \mathcal{N}(\beta_j, \sigma^2_{\beta_j}), j = 0, \ldots, k.$$

Least Squares Assumptions

Assumptions for causal inference:

- Causal inference is possible only under a strict set of assumptions.
  1. Errors have zero mean: $\mathsf{E}[\hat{u}_i | X_{1i} = x_1, \ldots, X_{ki} = x_k] = 0$. The conditional distribution of u_i given $X_{1i}, \ldots, X_{ki}$ has zero mean.
  2. Regressors are i.i.d.: $X_{1i}, \ldots, X_{ki} \sim$ i.i.d.. Random sampling implies the regressors are independently and identically distributed.
  3. Large outliers are unlikely: The regressand and regressors have finite kurtosis, $\mathsf{E}[Y^4] < \infty$, $\mathsf{E}[X_j^4] < \infty$.
  4. No perfect multicollinearity among the regressors.
- Under these assumptions, the least-squares estimators $\hat{\beta}_0, \ldots, \hat{\beta}_k$ are jointly normally distributed with

$$\hat{\beta}_j \sim \mathcal{N}(\beta_j, \sigma^2_{\hat{\beta}_j}), j = 0, \ldots, k.$$

Least Squares Assumptions

Assumptions for causal inference:

- Causal inference is possible only under a strict set of assumptions.
  1. Errors have zero mean: $\mathsf{E}[\hat{u}_i | X_{1i} = x_1, \ldots, X_{ki} = x_k] = 0$. The conditional distribution of u_i given $X_{1i}, \ldots, X_{ki}$ has zero mean.
  2. Regressors are i.i.d.: $X_{1i}, \ldots, X_{ki} \sim$ i.i.d.. Random sampling implies the regressors are independently and identically distributed.
  3. Large outliers are unlikely: The regressand and regressors have finite kurtosis, $\mathsf{E}[Y^4] < \infty, \mathsf{E}[X_j^4] < \infty$.
  4. No perfect multicollinearity among the regressors.
- Under these assumptions, the least-squares estimators $\hat{\beta}_0, \ldots, \hat{\beta}_k$ are jointly normally distributed with

$$\hat{\beta}_j \sim \mathcal{N}(\beta_j, \sigma^2_{\beta_j}), j = 0, \ldots, k.$$

Multicollinearity

Multicollinearity

The dummy variable trap:

- Consider binary variables that represent exhaustive and mutually exclusive states, e.g. true/-false; black/white; male/female; spring/summer/fall/winter; If there is an intercept in the regression, $\beta_0 \neq 0$, and if all binary variables are included as regressors, the regression will fail because of perfect multicollinearity.
- The usual way to avoid the dummy variable trap is to exclude one of the binary variables from the multiple regression, to eliminate the redundancy.
- Another way is to omit the intercept, that is estimate the coefficients $\beta_1, \ldots, \beta_k$ in the linear population regression without an intercept β_0:

$$\mathrm{E}[Y_i|X_{1i} = x_1, X_{2i} = x_2] = \beta_1 x_1 + \beta_2 x_2$$

Multicollinearity

The dummy variable trap:

- Consider binary variables that represent exhaustive and mutually exclusive states, e.g. true/-false; black/white; male/female; spring/summer/fall/winter; If there is an intercept in the regression, $\beta_0 \neq 0$, and if all binary variables are included as regressors, the regression will fail because of perfect multicollinearity.
- The usual way to avoid the dummy variable trap is to exclude one of the binary variables from the multiple regression, to eliminate the redundancy.
- Another way is to omit the intercept, that is estimate the coefficients $\beta_1, \ldots, \beta_k$ in the linear population regression without an intercept β_0:

$$\mathrm{E}[Y_i | X_{1i} = x_1, X_{2i} = x_2] = \beta_1 x_1 + \beta_2 x_2$$

Multicollinearity

The dummy variable trap:

- Consider binary variables that represent exhaustive and mutually exclusive states, e.g. true/-false; black/white; male/female; spring/summer/fall/winter; If there is an intercept in the regression, $\beta_0 \neq 0$, and if all binary variables are included as regressors, the regression will fail because of perfect multicollinearity.
- The usual way to avoid the dummy variable trap is to exclude one of the binary variables from the multiple regression, to eliminate the redundancy.
- Another way is to omit the intercept, that is estimate the coefficients $\beta_1, \ldots, \beta_k$ in the linear population regression without an intercept β_0:

$$\mathsf{E}[Y_i|X_{1i} = x_1, X_{2i} = x_2] = \beta_1 x_1 + \beta_2 x_2$$

Multicollinearity

Imperfect multicollinearity:

- Two or more of the regressors are highly correlated with each other.
- If the regressors are imperfectly multicollinear, then the coefficients on at least one individual regressor will be imprecisely estimated — they have a large sample variance.
- Let there be only two regressors and let the errors be homoskedastic. In this special case, the variance of the distribution reduces to:

  $$\sigma^2_{\hat{\beta}_1} = \frac{1}{n}\left(\frac{1}{1-\rho^2_{X_1X_2}}\right)\frac{\sigma^2_u}{\sigma^2_{X_1}}$$

  where $\rho_{X_1X_2}$ is the population correlation coefficient between the two regressors X_1 and X_2, and $\sigma^2_{X_1}$ is the population variance of X_1.
- If X_1 and X_2 are highly correlated, then $\rho^2_{X_1X_2} \approx 1$, so the term in the denominator is small and the variance of $\hat{\beta}_1$ is larger than it would be if $\rho_{X_1X_2} \approx 0$.
- Another feature is that $\hat{\beta}_1$ and $\hat{\beta}_2$ are, in general, correlated. With homoskedastic errors, the correlation may be computed as:

  $$\text{cor}(\hat{\beta}_1, \hat{\beta}_2) = -\rho_{X_1X_2}$$

Multicollinearity

Imperfect multicollinearity:

- Two or more of the regressors are highly correlated with each other.
- If the regressors are imperfectly multicollinear, then the coefficients on at least one individual regressor will be imprecisely estimated — they have a large sample variance.
- Let there be only two regressors and let the errors be homoskedastic. In this special case, the variance of the distribution reduces to:

$$\sigma^2_{\hat{\beta}_1} = \frac{1}{n}\left(\frac{1}{1-\rho^2_{X_1X_2}}\right)\frac{\sigma^2_u}{\sigma^2_{X_1}}$$

  where $\rho_{X_1X_2}$ is the population correlation coefficient between the two regressors X_1 and X_2, and $\sigma^2_{X_1}$ is the population variance of X_1.
- If X_1 and X_2 are highly correlated, then $\rho^2_{X_1X_2} \approx 1$, so the term in the denominator is small and the variance of $\hat{\beta}_1$ is larger than it would be if $\rho_{X_1X_2} \approx 0$.
- Another feature is that $\hat{\beta}_1$ and $\hat{\beta}_2$ are, in general, correlated. With homoskedastic errors, the correlation may be computed as:

$$\text{cor}(\hat{\beta}_1, \hat{\beta}_2) = -\rho_{X_1X_2}$$

Multicollinearity

Imperfect multicollinearity:

- Two or more of the regressors are highly correlated with each other.
- If the regressors are imperfectly multicollinear, then the coefficients on at least one individual regressor will be imprecisely estimated — they have a large sample variance.
- Let there be only two regressors and let the errors be homoskedastic. In this special case, the variance of the distribution reduces to:

$$\sigma^2_{\hat{\beta}_1} = \frac{1}{n}\left(\frac{1}{1-\rho^2_{X_1X_2}}\right)\frac{\sigma^2_u}{\sigma^2_{X_1}}$$

  where $\rho_{X_1X_2}$ is the population correlation coefficient between the two regressors X_1 and X_2, and $\sigma^2_{X_1}$ is the population variance of X_1.
- If X_1 and X_2 are highly correlated, then $\rho^2_{X_1X_2} \approx 1$, so the term in the denominator is small and the variance of $\hat{\beta}_1$ is larger than it would be if $\rho_{X_1X_2} \approx 0$.
- Another feature is that $\hat{\beta}_1$ and $\hat{\beta}_2$ are, in general, correlated. With homoskedastic errors, the correlation may be computed as:

$$\text{cor}(\hat{\beta}_1, \hat{\beta}_2) = -\rho_{X_1X_2}$$

Multicollinearity

Imperfect multicollinearity:

- Two or more of the regressors are highly correlated with each other.
- If the regressors are imperfectly multicollinear, then the coefficients on at least one individual regressor will be imprecisely estimated — they have a large sample variance.
- Let there be only two regressors and let the errors be homoskedastic. In this special case, the variance of the distribution reduces to:

$$\sigma^2_{\hat{\beta}_1} = \frac{1}{n}\left(\frac{1}{1-\rho^2_{X_1X_2}}\right)\frac{\sigma^2_u}{\sigma^2_{X_1}}$$

  where $\rho_{X_1X_2}$ is the population correlation coefficient between the two regressors X_1 and X_2, and $\sigma^2_{X_1}$ is the population variance of X_1.
- If X_1 and X_2 are highly correlated, then $\rho^2_{X_1X_2} \approx 1$, so the term in the denominator is small and the variance of $\hat{\beta}_1$ is larger than it would be if $\rho_{X_1X_2} \approx 0$.
- Another feature is that $\hat{\beta}_1$ and $\hat{\beta}_2$ are, in general, correlated. With homoskedastic errors, the correlation may be computed as:

$$\text{cor}(\hat{\beta}_1, \hat{\beta}_2) = -\rho_{X_1X_2}$$

Multicollinearity

Imperfect multicollinearity:

- Two or more of the regressors are highly correlated with each other.
- If the regressors are imperfectly multicollinear, then the coefficients on at least one individual regressor will be imprecisely estimated — they have a large sample variance.
- Let there be only two regressors and let the errors be homoskedastic. In this special case, the variance of the distribution reduces to:

$$\sigma^2_{\hat{\beta}_1} = \frac{1}{n}\left(\frac{1}{1-\rho^2_{X_1X_2}}\right)\frac{\sigma^2_u}{\sigma^2_{X_1}}$$

  where $\rho_{X_1X_2}$ is the population correlation coefficient between the two regressors X_1 and X_2, and $\sigma^2_{X_1}$ is the population variance of X_1.
- If X_1 and X_2 are highly correlated, then $\rho^2_{X_1X_2} \approx 1$, so the term in the denominator is small and the variance of $\hat{\beta}_1$ is larger than it would be if $\rho_{X_1X_2} \approx 0$.
- Another feature is that $\hat{\beta}_1$ and $\hat{\beta}_2$ are, in general, correlated. With homoskedastic errors, the correlation may be computed as:

$$\text{cor}(\hat{\beta}_1, \hat{\beta}_2) = -\rho_{X_1X_2}$$

Control Variables

Control Variables

Control Variables:

- A control variable is a regressor included to hold constant factors that, if neglected, could lead the estimated causal effect of interest to suffer from omitted variable bias.
- Effect of class size in California districts: Consider the potential omitted variable bias arising from omitting outside learning opportunities from a test score regression.
  - Outside learning opportunities is a broad concept that is difficult to measure.
  - Outside learning opportunities are correlated with the students' economic background, which can be measured.
- A measure of economic background can be included in a test score regression to control for omitted income-related determinants of test scores.

Control Variables

Control Variables:

- A control variable is a regressor included to hold constant factors that, if neglected, could lead the estimated causal effect of interest to suffer from omitted variable bias.
- Effect of class size in California districts: Consider the potential omitted variable bias arising from omitting outside learning opportunities from a test score regression.
  - Outside learning opportunities is a broad concept that is difficult to measure.
  - Outside learning opportunities are correlated with the students' economic background, which can be measured.
- A measure of economic background can be included in a test score regression to control for omitted income-related determinants of test scores.

Control Variables:

- A control variable is a regressor included to hold constant factors that, if neglected, could lead the estimated causal effect of interest to suffer from omitted variable bias.
- Effect of class size in California districts: Consider the potential omitted variable bias arising from omitting outside learning opportunities from a test score regression.
  - Outside learning opportunities is a broad concept that is difficult to measure.
  - Outside learning opportunities are correlated with the students' economic background, which can be measured.
- A measure of economic background can be included in a test score regression to control for omitted income-related determinants of test scores.

Control Variables

Control Variables:

- A control variable is a regressor included to hold constant factors that, if neglected, could lead the estimated causal effect of interest to suffer from omitted variable bias.
- Effect of class size in California districts: Consider the potential omitted variable bias arising from omitting outside learning opportunities from a test score regression.
  - Outside learning opportunities is a broad concept that is difficult to measure.
  - Outside learning opportunities are correlated with the students' economic background, which can be measured.
- A measure of economic background can be included in a test score regression to control for omitted income-related determinants of test scores.

Control Variables

Control Variables:

- A control variable is a regressor included to hold constant factors that, if neglected, could lead the estimated causal effect of interest to suffer from omitted variable bias.
- Effect of class size in California districts: Consider the potential omitted variable bias arising from omitting outside learning opportunities from a test score regression.
  - Outside learning opportunities is a broad concept that is difficult to measure.
  - Outside learning opportunities are correlated with the students' economic background, which can be measured.
- A measure of economic background can be included in a test score regression to control for omitted income-related determinants of test scores.

Control Variables

Class size in California districts:

- Augment the regression of test scores on STR and $PctEL$ with the percentage of students receiving a free or subsidized school lunch $LchPct$.
- Students are eligible for this program if their family income is less than a certain threshold (approximately 150% of the poverty line), so $LchPct$ measures the fraction of economically disadvantaged children in the district.
- The estimated regression is

$$\widehat{TestScore} = 700.2 - 1.00 \times STR - 0.122 \times PctEL - 0.547 \times LchPct$$

- Including the control variable $LchPct$ changes the coefficient on STR only slightly from -1.10 to -1.00.
- The coefficient on $LchPct$ is very large: The difference in test scores between a district with $LchPct = 0\%$ and one with $LchPct = 50\%$ is estimated to be 27.4 points, approximately the difference between the 75th and 25th percentiles of test scores.
- Would eliminating the reduced-price lunch program boost the district's test scores?
- If we treat the coefficient on STR as causal, why not the coefficient on $LchPct$?

Control Variables

Class size in California districts:

- Augment the regression of test scores on STR and $PctEL$ with the percentage of students receiving a free or subsidized school lunch $LchPct$.
- Students are eligible for this program if their family income is less than a certain threshold (approximately 150% of the poverty line), so $LchPct$ measures the fraction of economically disadvantaged children in the district.
- The estimated regression is

$$\widehat{TestScore} = 700.2 - 1.00 \times STR - 0.122 \times PctEL - 0.547 \times LchPct$$

- Including the control variable $LchPct$ changes the coefficient on STR only slightly from -1.10 to -1.00.
- The coefficient on $LchPct$ is very large: The difference in test scores between a district with $LchPct = 0\%$ and one with $LchPct = 50\%$ is estimated to be 27.4 points, approximately the difference between the 75th and 25th percentiles of test scores.
- Would eliminating the reduced-price lunch program boost the district's test scores?
- If we treat the coefficient on STR as causal, why not the coefficient on $LchPct$?

Control Variables

Class size in California districts:

- Augment the regression of test scores on STR and $PctEL$ with the percentage of students receiving a free or subsidized school lunch $LchPct$.
- Students are eligible for this program if their family income is less than a certain threshold (approximately 150% of the poverty line), so $LchPct$ measures the fraction of economically disadvantaged children in the district.
- The estimated regression is

$$\widehat{TestScore} = 700.2 - 1.00 \times STR - 0.122 \times PctEL - 0.547 \times LchPct$$

- Including the control variable $LchPct$ changes the coefficient on STR only slightly from -1.10 to -1.00.
- The coefficient on $LchPct$ is very large: The difference in test scores between a district with $LchPct = 0\%$ and one with $LchPct = 50\%$ is estimated to be 27.4 points, approximately the difference between the 75th and 25th percentiles of test scores.
- Would eliminating the reduced-price lunch program boost the district's test scores?
- If we treat the coefficient on STR as causal, why not the coefficient on $LchPct$?

Control Variables

Class size in California districts:

- Augment the regression of test scores on STR and $PctEL$ with the percentage of students receiving a free or subsidized school lunch $LchPct$.
- Students are eligible for this program if their family income is less than a certain threshold (approximately 150% of the poverty line), so $LchPct$ measures the fraction of economically disadvantaged children in the district.
- The estimated regression is

$$\widehat{TestScore} = 700.2 - 1.00 \times STR - 0.122 \times PctEL - 0.547 \times LchPct$$

- Including the control variable $LchPct$ changes the coefficient on STR only slightly from -1.10 to -1.00.
- The coefficient on $LchPct$ is very large: The difference in test scores between a district with $LchPct = 0\%$ and one with $LchPct = 50\%$ is estimated to be 27.4 points, approximately the difference between the 75th and 25th percentiles of test scores.
- Would eliminating the reduced-price lunch program boost the district's test scores?
- If we treat the coefficient on STR as causal, why not the coefficient on $LchPct$?

Control Variables

Class size in California districts:

- Augment the regression of test scores on STR and $PctEL$ with the percentage of students receiving a free or subsidized school lunch $LchPct$.
- Students are eligible for this program if their family income is less than a certain threshold (approximately 150% of the poverty line), so $LchPct$ measures the fraction of economically disadvantaged children in the district.
- The estimated regression is

$$\widehat{TestScore} = 700.2 - 1.00 \times STR - 0.122 \times PctEL - 0.547 \times LchPct$$

- Including the control variable $LchPct$ changes the coefficient on STR only slightly from -1.10 to -1.00.
- The coefficient on $LchPct$ is very large: The difference in test scores between a district with $LchPct = 0\%$ and one with $LchPct = 50\%$ is estimated to be 27.4 points, approximately the difference between the 75th and 25th percentiles of test scores.
- Would eliminating the reduced-price lunch program boost the district's test scores?
- If we treat the coefficient on STR as causal, why not the coefficient on $LchPct$?

Control Variables

Class size in California districts:

- Augment the regression of test scores on STR and $PctEL$ with the percentage of students receiving a free or subsidized school lunch $LchPct$.
- Students are eligible for this program if their family income is less than a certain threshold (approximately 150% of the poverty line), so $LchPct$ measures the fraction of economically disadvantaged children in the district.
- The estimated regression is

$$\widehat{TestScore} = 700.2 - 1.00 \times STR - 0.122 \times PctEL - 0.547 \times LchPct$$

- Including the control variable $LchPct$ changes the coefficient on STR only slightly from -1.10 to -1.00.
- The coefficient on $LchPct$ is very large: The difference in test scores between a district with $LchPct = 0\%$ and one with $LchPct = 50\%$ is estimated to be 27.4 points, approximately the difference between the 75th and 25th percentiles of test scores.
- Would eliminating the reduced-price lunch program boost the district's test scores?
- If we treat the coefficient on STR as causal, why not the coefficient on $LchPct$?

Control Variables

Class size in California districts:

- Augment the regression of test scores on STR and $PctEL$ with the percentage of students receiving a free or subsidized school lunch $LchPct$.
- Students are eligible for this program if their family income is less than a certain threshold (approximately 150% of the poverty line), so $LchPct$ measures the fraction of economically disadvantaged children in the district.
- The estimated regression is

$$\widehat{TestScore} = 700.2 - 1.00 \times STR - 0.122 \times PctEL - 0.547 \times LchPct$$

- Including the control variable $LchPct$ changes the coefficient on STR only slightly from -1.10 to -1.00.
- The coefficient on $LchPct$ is very large: The difference in test scores between a district with $LchPct = 0\%$ and one with $LchPct = 50\%$ is estimated to be 27.4 points, approximately the difference between the 75th and 25th percentiles of test scores.
- Would eliminating the reduced-price lunch program boost the district's test scores?
- If we treat the coefficient on STR as causal, why not the coefficient on $LchPct$?

Conditional Mean Independence

Conditional Mean Independence:

- The conditional expectation of the error terms (given the variables of interest and controls) is independent of the variable of interest (but it can depend on the control variables).
- By including control variables, the variables of interest are no longer correlated with the error term. If conditional mean independence holds, the regressors of interest can be treated as if they were randomly assigned: After adding control variables in the regression, the conditional mean of the error term is independent of the regressors.
- The least-squares estimators of the coefficients on the X_is are unbiased estimators of the causal effects of the X_is. The least-squares estimators of the coefficients on the W_is are biased, but their estimates are of no special interest.
- **Class size in California districts:**
  $LchPct$ is correlated with factors that enter the error term, such as learning opportunities outside school, and does not have a causal interpretation. If the conditional mean independence assumption holds, the mean of the error term, given the control variables $PctEL$ and $LchPct$, does not depend on the student-teacher ratio. Thus, among schools with the same values of $PctEL$ and $LchPct$, class size is "as-if" randomly assigned.

Conditional Mean Independence

Conditional Mean Independence:

- The conditional expectation of the error terms (given the variables of interest and controls) is independent of the variable of interest (but it can depend on the control variables).
- By including control variables, the variables of interest are no longer correlated with the error term. If conditional mean independence holds, the regressors of interest can be treated as if they were randomly assigned: After adding control variables in the regression, the conditional mean of the error term is independent of the regressors.
- The least-squares estimators of the coefficients on the X_is are unbiased estimators of the causal effects of the X_is. The least-squares estimators of the coefficients on the W_is are biased, but their estimates are of no special interest.
- **Class size in California districts:**
  $LchPct$ is correlated with factors that enter the error term, such as learning opportunities outside school, and does not have a causal interpretation. If the conditional mean independence assumption holds, the mean of the error term, given the control variables $PctEL$ and $LchPct$, does not depend on the student-teacher ratio. Thus, among schools with the same values of $PctEL$ and $LchPct$, class size is "as-if" randomly assigned.

Conditional Mean Independence

Conditional Mean Independence:

- The conditional expectation of the error terms (given the variables of interest and controls) is independent of the variable of interest (but it can depend on the control variables).
- By including control variables, the variables of interest are no longer correlated with the error term. If conditional mean independence holds, the regressors of interest can be treated as if they were randomly assigned: After adding control variables in the regression, the conditional mean of the error term is independent of the regressors.
- The least-squares estimators of the coefficients on the X_is are unbiased estimators of the causal effects of the X_is. The least-squares estimators of the coefficients on the W_is are biased, but their estimates are of no special interest.
- **Class size in California districts:** $LchPct$ is correlated with factors that enter the error term, such as learning opportunities outside school, and does not have a causal interpretation. If the conditional mean independence assumption holds, the mean of the error term, given the control variables $PctEL$ and $LchPct$, does not depend on the student-teacher ratio. Thus, among schools with the same values of $PctEL$ and $LchPct$, class size is "as-if" randomly assigned.

Conditional Mean Independence

Conditional Mean Independence:

- The conditional expectation of the error terms (given the variables of interest and controls) is independent of the variable of interest (but it can depend on the control variables).
- By including control variables, the variables of interest are no longer correlated with the error term. If conditional mean independence holds, the regressors of interest can be treated as if they were randomly assigned: After adding control variables in the regression, the conditional mean of the error term is independent of the regressors.
- The least-squares estimators of the coefficients on the X_is are unbiased estimators of the causal effects of the X_is. The least-squares estimators of the coefficients on the W_is are biased, but their estimates are of no special interest.
- **Class size in California districts:**
  $LchPct$ is correlated with factors that enter the error term, such as learning opportunities outside school, and does not have a causal interpretation. If the conditional mean independence assumption holds, the mean of the error term, given the control variables $PctEL$ and $LchPct$, does not depend on the student-teacher ratio. Thus, among schools with the same values of $PctEL$ and $LchPct$, class size is “as-if” randomly assigned.

Summary

Summary

- Omitted variable bias occurs when an omitted variable (a) is correlated with an included regressor and (b) is a determinant of Y.
- The multiple regression model is a linear regression model that includes multiple regressors, $X_1, X_2, \ldots, X_k$. Associated with each regressor is a regression coefficient, $\beta_1, \beta_2, \ldots, \beta_k$. The coefficient β_1 is the expected difference in Y associated with a one-unit difference in X_1, holding the other regressors constant.
- The coefficients in multiple regression can be estimated by OLS. When the Gauss-Markov assumptions hold, the OLS estimators of the causal effect are unbiased, consistent, and normally distributed in large samples.
- The role of control variables is to hold constant omitted factors so that the variable of interest is no longer correlated with the error term.
- Perfect multicollinearity occurs when one regressor is an exact linear function of the other regressors.
- The standard error of the regression, the R^2, and the $\bar{R}^2$ are measures of fit for the multiple regression model.

Summary

- Omitted variable bias occurs when an omitted variable (a) is correlated with an included regressor and (b) is a determinant of Y.
- The multiple regression model is a linear regression model that includes multiple regressors, $X_1, X_2, \ldots, X_k$. Associated with each regressor is a regression coefficient, $\beta_1, \beta_2, \ldots, \beta_k$. The coefficient β_1 is the expected difference in Y associated with a one-unit difference in X_1, holding the other regressors constant.
- The coefficients in multiple regression can be estimated by OLS. When the Gauss-Markov assumptions hold, the OLS estimators of the causal effect are unbiased, consistent, and normally distributed in large samples.
- The role of control variables is to hold constant omitted factors so that the variable of interest is no longer correlated with the error term.
- Perfect multicollinearity occurs when one regressor is an exact linear function of the other regressors.
- The standard error of the regression, the R^2, and the $\bar{R}^2$ are measures of fit for the multiple regression model.

Summary

- Omitted variable bias occurs when an omitted variable (a) is correlated with an included regressor and (b) is a determinant of Y.
- The multiple regression model is a linear regression model that includes multiple regressors, $X_1, X_2, \ldots, X_k$. Associated with each regressor is a regression coefficient, $\beta_1, \beta_2, \ldots, \beta_k$. The coefficient β_1 is the expected difference in Y associated with a one-unit difference in X_1, holding the other regressors constant.
- The coefficients in multiple regression can be estimated by OLS. When the Gauss-Markov assumptions hold, the OLS estimators of the causal effect are unbiased, consistent, and normally distributed in large samples.
- The role of control variables is to hold constant omitted factors so that the variable of interest is no longer correlated with the error term.
- Perfect multicollinearity occurs when one regressor is an exact linear function of the other regressors.
- The standard error of the regression, the R^2, and the $\bar{R}^2$ are measures of fit for the multiple regression model.

Summary

- Omitted variable bias occurs when an omitted variable (a) is correlated with an included regressor and (b) is a determinant of Y.
- The multiple regression model is a linear regression model that includes multiple regressors, $X_1, X_2, \ldots, X_k$. Associated with each regressor is a regression coefficient, $\beta_1, \beta_2, \ldots, \beta_k$. The coefficient β_1 is the expected difference in Y associated with a one-unit difference in X_1, holding the other regressors constant.
- The coefficients in multiple regression can be estimated by OLS. When the Gauss-Markov assumptions hold, the OLS estimators of the causal effect are unbiased, consistent, and normally distributed in large samples.
- The role of control variables is to hold constant omitted factors so that the variable of interest is no longer correlated with the error term.
- Perfect multicollinearity occurs when one regressor is an exact linear function of the other regressors.
- The standard error of the regression, the R^2, and the $\bar{R}^2$ are measures of fit for the multiple regression model.

Summary

- Omitted variable bias occurs when an omitted variable (a) is correlated with an included regressor and (b) is a determinant of Y.
- The multiple regression model is a linear regression model that includes multiple regressors, $X_1, X_2, \ldots, X_k$. Associated with each regressor is a regression coefficient, $\beta_1, \beta_2, \ldots, \beta_k$. The coefficient β_1 is the expected difference in Y associated with a one-unit difference in X_1, holding the other regressors constant.
- The coefficients in multiple regression can be estimated by OLS. When the Gauss-Markov assumptions hold, the OLS estimators of the causal effect are unbiased, consistent, and normally distributed in large samples.
- The role of control variables is to hold constant omitted factors so that the variable of interest is no longer correlated with the error term.
- Perfect multicollinearity occurs when one regressor is an exact linear function of the other regressors.
- The standard error of the regression, the R^2, and the $\bar{R}^2$ are measures of fit for the multiple regression model.

Summary

- Omitted variable bias occurs when an omitted variable (a) is correlated with an included regressor and (b) is a determinant of Y.
- The multiple regression model is a linear regression model that includes multiple regressors, $X_1, X_2, \ldots, X_k$. Associated with each regressor is a regression coefficient, $\beta_1, \beta_2, \ldots, \beta_k$. The coefficient β_1 is the expected difference in Y associated with a one-unit difference in X_1, holding the other regressors constant.
- The coefficients in multiple regression can be estimated by OLS. When the Gauss-Markov assumptions hold, the OLS estimators of the causal effect are unbiased, consistent, and normally distributed in large samples.
- The role of control variables is to hold constant omitted factors so that the variable of interest is no longer correlated with the error term.
- Perfect multicollinearity occurs when one regressor is an exact linear function of the other regressors.
- The standard error of the regression, the R^2, and the $\bar{R}^2$ are measures of fit for the multiple regression model.

Problems & Applications

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 6, Review Question 3.

How does $\bar{R}^2$ differ from R^2? Why is $\bar{R}^2$ useful in a regression model with multiple regressors?

Stock & Watson, Introduction (4th), Chapter 6, Review Question 6.

Explain why it is difficult to estimate precisely the partial effect of X_1, holding X_2 constant, if X_1 and X_2 are highly correlated.

Stock & Watson, Introduction (4th), Chapter 6, Exercise 4.

Consider the regression of average hourly earnings AHE (in dollars) on Age (in years) and several binary variables for characteristics such as sex, education, and region of employment:

$$\begin{aligned}\widehat{AHE} = 0.33 &+ 10.42\,College - 4.57\,Female + 0.61\,Age \\ &+ 0.74\,Northeast - 1.54\,Midwest - 0.44\,South\end{aligned}$$

$$R^2 = 0.185, \quad SER = 12.01, \quad n = 7178$$

1. Do there appear to be important regional differences?
2. Why is the regressor $West$ omitted from the regression? What would happen if it were included?
3. Juanita is a 28-year-old female college graduate from the South. Jennifer is a 28-year-old female college graduate from the Midwest. Calculate the expected difference in earnings between Juanita and Jennifer.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 6, Exercise 4.

Consider the regression of average hourly earnings AHE (in dollars) on Age (in years) and several binary variables for characteristics such as sex, education, and region of employment:

$$\begin{aligned}\widehat{AHE} = 0.33 + 10.42\, College - 4.57\, Female + 0.61\, Age \\ + 0.74\, Northeast - 1.54\, Midwest - 0.44\, South\end{aligned}$$

$$R^2 = 0.185, \quad SER = 12.01, \quad n = 7178$$

1. Do there appear to be important regional differences?
2. Why is the regressor $West$ omitted from the regression? What would happen if it were included?
3. Juanita is a 28-year-old female college graduate from the South. Jennifer is a 28-year-old female college graduate from the Midwest. Calculate the expected difference in earnings between Juanita and Jennifer.

Stock & Watson, Introduction (4th), Chapter 6, Exercise 4.

Consider the regression of average hourly earnings AHE (in dollars) on Age (in years) and several binary variables for characteristics such as sex, education, and region of employment:

$$\begin{aligned}\widehat{AHE} = 0.33 + 10.42\,College - 4.57\,Female + 0.61\,Age \\ + 0.74\,Northeast - 1.54\,Midwest - 0.44\,South\end{aligned}$$

$$R^2 = 0.185, \quad SER = 12.01, \quad n = 7178$$

1. Do there appear to be important regional differences?
2. Why is the regressor $West$ omitted from the regression? What would happen if it were included?
3. Juanita is a 28-year-old female college graduate from the South. Jennifer is a 28-year-old female college graduate from the Midwest. Calculate the expected difference in earnings between Juanita and Jennifer.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 6, Exercise 5.

Data were collected from a random sample of 220 home sales from a community in 2013. Let $Price$ denote the selling price (in $\$1000$s), BDR denote the number of bedrooms, $Bath$ denote the number of bathrooms, $Hsize$ denote the size of the house (in square feet), $Lsize$ denote the lot size (in square feet), Age denote the age of the house (in years), and $Poor$ denote a binary variable that is equal to 1 if the condition of the house is reported as "poor." An estimated regression yields

$$\begin{aligned}\widehat{Price} =& 119.2 + 0.485\,BDR + 23.4\,Bath + 0.156\,Hsize \\ &+ 0.002\,Lsize + 0.090\,Age - 48.8\,Poor \quad \bar{R}^2 = 0.72, \quad SER = 41.5\end{aligned}$$

1. Suppose a homeowner converts part of an existing family room in her house into a new bathroom. What is the expected increase in the value of the house?
2. Suppose a homeowner adds a new bathroom to her house, which increases the size of the house by 100 square feet. What is the expected increase in the value of the house?
3. What is the loss in value if a homeowner lets his house run down, so that its condition becomes "poor"?
4. Compute the R^2 for the regression.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 6, Exercise 5.

Data were collected from a random sample of 220 home sales from a community in 2013. Let $Price$ denote the selling price (in $\$1000$s), BDR denote the number of bedrooms, $Bath$ denote the number of bathrooms, $Hsize$ denote the size of the house (in square feet), $Lsize$ denote the lot size (in square feet), Age denote the age of the house (in years), and $Poor$ denote a binary variable that is equal to 1 if the condition of the house is reported as "poor." An estimated regression yields

$$\begin{aligned}\widehat{Price} =& 119.2 + 0.485\, BDR + 23.4\, Bath + 0.156\, Hsize \\ &+ 0.002\, Lsize + 0.090\, Age - 48.8\, Poor \quad \bar{R}^2 = 0.72, \quad SER = 41.5\end{aligned}$$

1. Suppose a homeowner converts part of an existing family room in her house into a new bathroom. What is the expected increase in the value of the house?
2. Suppose a homeowner adds a new bathroom to her house, which increases the size of the house by 100 square feet. What is the expected increase in the value of the house?
3. What is the loss in value if a homeowner lets his house run down, so that its condition becomes "poor"?
4. Compute the R^2 for the regression.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 6, Exercise 5.

Data were collected from a random sample of 220 home sales from a community in 2013. Let $Price$ denote the selling price (in $\$1000$s), BDR denote the number of bedrooms, $Bath$ denote the number of bathrooms, $Hsize$ denote the size of the house (in square feet), $Lsize$ denote the lot size (in square feet), Age denote the age of the house (in years), and $Poor$ denote a binary variable that is equal to 1 if the condition of the house is reported as "poor." An estimated regression yields

$$\begin{aligned}\widehat{Price} =& 119.2 + 0.485\, BDR + 23.4\, Bath + 0.156\, Hsize \\ & + 0.002\, Lsize + 0.090\, Age - 48.8\, Poor \quad \bar{R}^2 = 0.72, \quad SER = 41.5\end{aligned}$$

1. Suppose a homeowner converts part of an existing family room in her house into a new bathroom. What is the expected increase in the value of the house?
2. Suppose a homeowner adds a new bathroom to her house, which increases the size of the house by 100 square feet. What is the expected increase in the value of the house?
3. What is the loss in value if a homeowner lets his house run down, so that its condition becomes "poor"?
4. Compute the R^2 for the regression.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 6, Exercise 5.

Data were collected from a random sample of 220 home sales from a community in 2013. Let $Price$ denote the selling price (in $\$1000$s), BDR denote the number of bedrooms, $Bath$ denote the number of bathrooms, $Hsize$ denote the size of the house (in square feet), $Lsize$ denote the lot size (in square feet), Age denote the age of the house (in years), and $Poor$ denote a binary variable that is equal to 1 if the condition of the house is reported as “poor.” An estimated regression yields

$$\begin{aligned}\widehat{Price} =& 119.2 + 0.485\, BDR + 23.4\, Bath + 0.156\, Hsize \\ &+ 0.002\, Lsize + 0.090\, Age - 48.8\, Poor \quad \bar{R}^2 = 0.72, \quad SER = 41.5\end{aligned}$$

1. Suppose a homeowner converts part of an existing family room in her house into a new bathroom. What is the expected increase in the value of the house?
2. Suppose a homeowner adds a new bathroom to her house, which increases the size of the house by 100 square feet. What is the expected increase in the value of the house?
3. What is the loss in value if a homeowner lets his house run down, so that its condition becomes “poor”?
4. Compute the R^2 for the regression.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 6, Exercise 6.

A researcher plans to study the causal effect of police on crime, using data from a random sample of U.S. counties. He plans to regress the county's crime rate on the (per capita) size of the county's police force.

1. Explain why this regression is likely to suffer from omitted variable bias. Which variables would you add to the regression to control for important omitted variables?
2. Use the previous answer to determine whether the regression will likely over- or underestimate the effect of police on the crime rate. That is, do you think that $\hat{\beta}_1 > \beta_1$ or $\hat{\beta}_1 < \beta_1$?

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 6, Exercise 6.

A researcher plans to study the causal effect of police on crime, using data from a random sample of U.S. counties. He plans to regress the county's crime rate on the (per capita) size of the county's police force.

1. Explain why this regression is likely to suffer from omitted variable bias. Which variables would you add to the regression to control for important omitted variables?
2. Use the previous answer to determine whether the regression will likely over- or underestimate the effect of police on the crime rate. That is, do you think that $\hat{\beta}_1 > \beta_1$ or $\hat{\beta}_1 < \beta_1$?