Assessing Studies Based on Multiple Regression

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.

In this lesson you will learn ...

- Internal and external validity.
- Functional form misspecification.
- Errors-in-variables bias.
- Sample selection bias.
- Simultaneous causality bias.

Internal and External Validity

Internal and external validity

- **Definition:** Distinguish between the population and setting from which the sample was drawn and the population and setting to which the causal inferences from the study are to be applied. **Internal validity:** The statistical inferences about causal effects are valid for the population under study. **External validity:** The statistical inferences about causal effects can be generalized from the population and setting studied to other populations and settings.
- **Threats to external validity:**
  1. **Differences in populations:** The selected population may differ from the population of interest, e.g. geographical differences, individual differences.
  2. **Differences in settings:** The legal setting in which the study was conducted may differ from the legal setting to which its results are applied.
- **Threats to internal validity:**
  1. **Omitted variable bias**
  2. **Functional form misspecification**
  3. **Errors-in-variables bias**
  4. **Sample selection**
  5. **Simultaneous causality**

Threats to Internal Validity

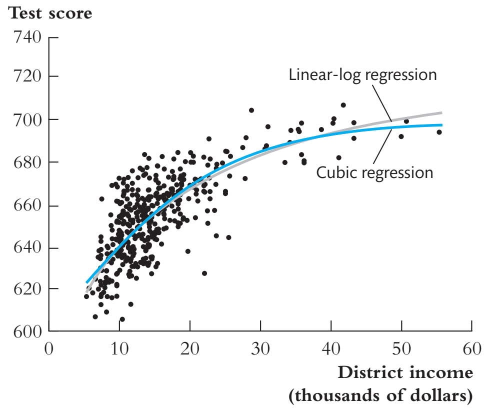
1. **Omitted Variable Bias:** Adding a variable to a regression has both costs and benefits: Including the variable when it does not belong reduces the precision of the estimators of the other regression coefficients. Trade-off: **bias vs. variance** of the coefficient of interest.
2. **Functional Form Misspecification:** The functional form of the estimated regression function differs from that of the population regression function. The estimator of the partial effect of a change in one of the variables could be biased.
3. **Errors-in-Variables Bias:** The effect of measurement error in Y is different from that of measurement error in X. If an explanatory variable X is measured imprecisely, the bias persists in large samples. If Y is measured imprecisely, the variance of $\hat{\beta}_1$ is increased, but $\hat{\beta}_1$ is not biased. Solutions: Instrumental variables regression; develop a mathematical model of the measurement error.
4. **Sample Selection:** If the selection process influences the availability of data and that process is related to the dependent variable beyond depending on the regressors. Such sample selection induces correlation between one or more regressors and the error term, leading to bias and inconsistency of the OLS estimator.
5. **Simultaneous Causality:** If causality also runs from the dependent variable to one or more regressors (Y causes X), the OLS estimator is biased and inconsistent. Solutions: instrumental variables regression; randomized controlled experiment in which the reverse causality channel is neutralized.

External Validity: California and Massachusetts

- **External validity:** Does California data generalize to other standardized tests in other elementary public school districts in the United States?
- We examine a different data set, based on standardized test results for fourth graders in 220 public school districts in Massachusetts in 1998. Both the Massachusetts and California tests are broad measures of student knowledge and academic skills, although the details differ. The organization of classroom instruction is broadly similar at the elementary school level in the two states, but aspects of elementary school funding and curriculum differ.
- The average test score is higher in Massachusetts, but the test is different, so a direct comparison of scores is not appropriate.
- The average student-teacher ratio is higher in California: 19.6 vs. 17.3.
- Average district income is 20% higher in Massachusetts, but the standard deviation of district income is greater in California.
- The average percentage of students still learning English and the average percentage of students receiving subsidized lunches are both much higher in California.

Test Scores and Class Size: Control for Average District Income

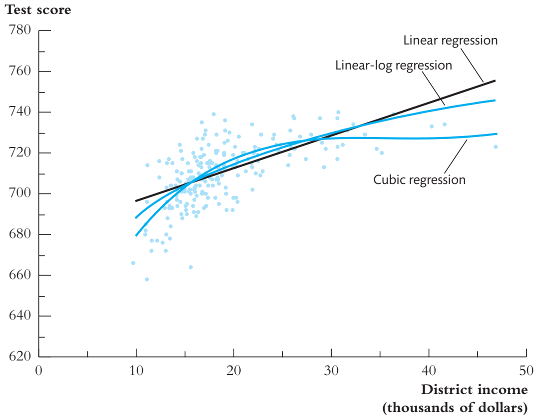
Test Scores vs District Income in California



The estimated linear-log and cubic regression functions are nearly identical in the California districts sample.

Test Scores and Class Size: Control for Average District Income

Test Scores vs District Income in Massachusetts



The estimated linear-log and cubic regression functions are similar for district incomes between $13,000 and $30,000, the region containing most of the observations.

Test Scores and Class Size

Comparison of the California and Massachusetts data

	California		Massachusetts	
	Average	Standard Deviation	Average	Standard Deviation
Test scores	654.1	19.1	709.8	15.1
Student–teacher ratio	19.6	1.9	17.3	2.3
% English learners	15.8%	18.3%	1.1%	2.9%
% receiving subsidized lunch	44.7%	27.1%	15.3%	15.1%
Average district income ($)	$15,317	$7226	$18,747	$5808
Number of observations	420		220	
Year	1999		1998	

Regressor	(1)	(2)	(3)	(4)	(5)	(6)
Student–teacher ratio (STR)	−1.72 (0.50) [−2.70, −0.73]	−0.69 (0.27) [−1.22, −0.16]	−0.64 (0.27) [−1.17, −0.11]	12.4 (14.0)	−1.02 (0.37)	−0.67 (0.27) [−1.21, −0.14]
STR^2				−0.680 (0.737)		
STR^3				0.011 (0.013)		
% English learners		−0.411 (0.306)	−0.437 (0.303)	−0.434 (0.300)		
% English learners > median? (Binary, $HiEL$)					−12.6 (9.8)	
$HiEL \times STR$					0.80 (0.56)	
% eligible for free lunch		−0.521 (0.077)	−0.582 (0.097)	−0.587 (0.104)	−0.709 (0.091)	−0.653 (0.72)
District income (logarithm)		16.53 (3.15)				
District income			−3.07 (2.35)	−3.38 (2.49)	−3.87 (2.49)	−3.22 (2.31)
District income^2			0.164 (0.085)	0.174 (0.089)	0.184 (0.090)	0.165 (0.085)
District income^3			−0.0022 (0.0010)	−0.0023 (0.0010)	−0.0023 (0.0010)	−0.0022 (0.0010)
***F*-Statistics and *p*-Values Testing Exclusion of Groups of Variables**						
All STR variables and interactions $= 0$				2.86 (0.038)	4.01 (0.020)	
$STR^2, STR^3 = 0$				0.45 (0.641)		
$Income^2, Income^3$			7.74 (< 0.001)	7.75 (< 0.001)	5.85 (0.003)	6.55 (0.002)
$HiEL, HiEL \times STR$					1.58 (0.208)	
SER	14.64	8.69	8.61	8.63	8.62	8.64
$\overline{R}^2$	0.063	0.670	0.676	0.675	0.675	0.674

Test Scores and Class Size

Main findings for California

1. Adding variables that control for student background characteristics reduces the coefficient on the student-teacher ratio from -2.28 to -0.73.
2. The hypothesis that the true coefficient on the student-teacher ratio is 0 is rejected at the 1% significance level, even after adding variables that control for student background and district economic characteristics.
3. The effect of cutting the student-teacher ratio does not depend in a statistically significant way on the percentage of English learners in the district.
4. There is some evidence that the relationship between test scores and the student-teacher ratio is nonlinear.

Do we these findings carry over to Massachusetts? Yes for (1)-(3), but no for (4).

Test Scores and Class Size

Main findings for Massachusetts

1. Adding variables that control for student background characteristics reduces the coefficient on the student-teacher ratio from -1.72 to -0.69.
2. The hypothesis that the true coefficient on the student-teacher ratio is 0 is rejected at the 5% significance level, whereas it is rejected at the 1% level in the California data. However, the California sample is much larger, so the California estimates are more precise.
3. The effect of cutting the student-teacher ratio does not depend in a statistically significant way on the percentage of English learners in the district.
4. The hypothesis that the relationship between the student-teacher ratio and test scores is linear cannot be rejected at the 5% significance level when tested against a cubic specification.

Test Scores and Class Size

Comparing California and Massachusetts

- The standardized tests are different — One point on the Massachusetts test is not the same as one point on the California test — the regression coefficients cannot be compared directly.
- Standardize the test scores: Subtract the sample average and divide by the standard deviation so that they have a mean of 0 and a variance of 1. The slope coefficients in the regression with the standardized test score equal the slope coefficients in the original regression divided by the standard deviation of the test.
- The coefficient on the student-teacher ratio divided by the standard deviation of test scores can be compared across the two data sets.
- The OLS coefficient estimate using California data is -0.73, so cutting the student-teacher ratio by two is estimated to increase district test scores by $-0.73 \times (-2) = 1.46$ points. The standard deviation of test scores is 19.1 points, so the standardized gain is $1.46/19.1 = 0.076$ standard deviation units of the distribution of test scores across districts. The standard error of this estimate is $0.26 \times 2/19.1 = 0.027$.

Test Scores and Class Size

Massachusetts data suggests California results are externally valid

- Based on the linear model using California data, a reduction of two students per teacher is estimated to increase test scores by 0.076 standard deviation units, with a standard error of 0.027.
- The nonlinear models for California data suggest a somewhat larger effect, with the specific effect depending on the initial student-teacher ratio. Based on the Massachusetts data, this estimated effect is 0.085 standard deviation units, with a standard error of 0.036.
- The 95% confidence interval for Massachusetts contains the 95% confidence interval for the California linear specification.
- Cutting the student-teacher ratio is predicted to raise test scores, but the predicted improvement is small.

Test Scores and Class Size

Addressing potential threats to internal validity

- **Omitted variables:** Possibly omitted: teacher quality, e.g. better teachers are attracted to schools with smaller student-teacher ratios; districts with low teacher-student ratios attract families more committed to achieving high test scores. Solution: design an experiment with random assignment of students.
- **Functional form:** Introducing nonlinearities does not substantially alter estimates of the coefficient on the student-teacher ratio.
- **Errors in variables:** Average district income is from the 1990 Census, while the other data pertain to 1998 (Massachusetts) or 1999 (California). If the economic composition of the district changed substantially over the 1990s, this would be an imprecise measure of the actual average district income.
- **Sample selection:** The sample is exhaustive as it covers all public elementary school districts in the state that satisfy minimum size restrictions.
- **Simultaneous causality:** Would arise if the performance on standardized tests affected the student-teacher ratio. This could happen if the funding of poorly performing schools or districts resulted in more teachers being hired. In California, court cases led to some equalization of funding, but this redistribution of funds was not based on student achievement. In Massachusetts no such mechanism was in place.
- **Internal validity:** The study controls for student background, family economic background, and district affluence; and checks for non-linearities in the regression function.

Summary

1. A study is internally valid if the statistical inferences about causal effects are valid for the population being studied.
2. A study is externally valid if its inferences and conclusions can be generalized from the population and setting studied to other populations and settings.
3. There are two types of threats to internal validity: 1. OLS estimators are biased and inconsistent if the regressors and error terms are correlated. 2. Confidence intervals and hypothesis tests are not valid when the standard errors are incorrect.
4. Regressors and error terms may be correlated when there are omitted variables, an incorrect functional form is used, one or more of the regressors are measured with error, the sample is chosen non-randomly from the population, or there is simultaneous causality between the regressors and dependent variables.
5. Standard errors are incorrect when the error term is correlated across different observations. Stardard errors could be incorrect if the errors are heteroskedastic and the computer software does not use robust estimates.
6. When regression models are used solely for prediction, it is not necessary for the regression coefficients to be unbiased estimates of causal effects, but the regression model be externally valid.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 9, Exercise 1.

Suppose you have just read a careful statistical study of the effect of advertising on the demand for cigarettes. Using data from New York during the 1970s, the study concluded that advertising on buses and subways was more effective than print advertising. Use the concept of external validity to determine if these results are likely to apply to Boston in the 1970s, Los Angeles in the 1970s, and New York in 2018.

Stock & Watson, Introduction (4th), Chapter 9, Exercise 3.

Labor economists studying the determinants of women's earnings discovered a puzzling empirical result. Using randomly selected employed women, they regressed earnings on the women's number of children and a set of control variables (age, education, occupation, and so forth). They found that women with more children had higher wages, controlling for these other factors. Explain how sample selection might be the cause of this result.