Regression Review: Economics and Public Policy

Brian Knight, Brown University

January 28, 2010

Introduction

- Suppose that one wants to measure the effects of some policy, say charter school attendance, on an individual outcome, say test scores.

Introduction

- Suppose that one wants to measure the effects of some policy, say charter school attendance, on an individual outcome, say test scores.
- One way to measure this relationship is through a regression model:

$$Y_i = \alpha + \beta X_i + u_i$$

Introduction

- Suppose that one wants to measure the effects of some policy, say charter school attendance, on an individual outcome, say test scores.
- One way to measure this relationship is through a regression model:

$$Y_i = \alpha + \beta X_i + u_i$$

- where Y_i is the test score of individual i, X_i is charter school attendance, and u_i is unmeasured ability.

Introduction

- The effect of charter school attendance, *holding ability constant*, is given by β.

Introduction

- The effect of charter school attendance, *holding ability constant*, is given by β.
- This effect is typically measured via an OLS regression.

Introduction

- The effect of charter school attendance, *holding ability constant*, is given by β.
- This effect is typically measured via an OLS regression.
- The standard OLS estimate of β is given by:

$$\widehat{\beta}_{OLS} = \frac{cov(Y_i, X_i)}{var(X_i)} = \beta + \frac{cov(X_i, u_i)}{var(X_i)}$$

- The effect of charter school attendance, *holding ability constant*, is given by β.
- This effect is typically measured via an OLS regression.
- The standard OLS estimate of β is given by:

$$\widehat{\beta}_{OLS} = \frac{cov(Y_i, X_i)}{var(X_i)} = \beta + \frac{cov(X_i, u_i)}{var(X_i)}$$

- Thus, the estimated effect of charter schools captures the true effect of charter schools only if attendance is uncorrelated with ability $[cov(X_i, u_i) = 0]$. Otherwise, the measured effect may suffer from "selection" or "endogeneity" problems.

Correction 1: Controlling for observables

- One approach to correcting endogeneity problems is to attempt to measure ability through multiple regression analysis:

$$Y_i = \alpha + \beta X_i + \gamma W_i + u_i$$

Correction 1: Controlling for observables

- One approach to correcting endogeneity problems is to attempt to measure ability through multiple regression analysis:

$$Y_i = \alpha + \beta X_i + \gamma W_i + u_i$$

- Unfortunately, the researcher often does not have access to comprehensive measures of ability. And other factors, such as motivation, are even harder to measure. So, multiple regression helps to reduce the role of unobservables but typically does not solve the problem.

Correction 2: Instrumental Variables

- The ideal experiment would be randomization (as in clinical trials). Unfortunately, such experiments are not always feasible.

Correction 2: Instrumental Variables

- The ideal experiment would be randomization (as in clinical trials). Unfortunately, such experiments are not always feasible.
- Researchers thus look to "natural experiments". In particular, the researcher finds an instrument (Z_i) that is uncorrelated with ability but that is correlated with charter school attendance. For example, some charter schools are over-subscribed and attendance is determined via a lottery.

$$\widehat{\beta}_{IV} = \frac{cov(Y_i, Z_i)}{cov(Z_i, X_i)} = \beta + \frac{cov(Z_i, u_i)}{cov(Z_i, X_i)}$$

Correction 2: Instrumental Variables

- The ideal experiment would be randomization (as in clinical trials). Unfortunately, such experiments are not always feasible.
- Researchers thus look to "natural experiments". In particular, the researcher finds an instrument (Z_i) that is uncorrelated with ability but that is correlated with charter school attendance. For example, some charter schools are over-subscribed and attendance is determined via a lottery.

$$\widehat{\beta}_{IV} = \frac{cov(Y_i, Z_i)}{cov(Z_i, X_i)} = \beta + \frac{cov(Z_i, u_i)}{cov(Z_i, X_i)}$$

- One way to implement instrumental variables estimation is via two-stage least squares.

Correction 2: Instrumental Variables

- The ideal experiment would be randomization (as in clinical trials). Unfortunately, such experiments are not always feasible.
- Researchers thus look to "natural experiments". In particular, the researcher finds an instrument (Z_i) that is uncorrelated with ability but that is correlated with charter school attendance. For example, some charter schools are over-subscribed and attendance is determined via a lottery.

$$\widehat{\beta}_{IV} = \frac{cov(Y_i, Z_i)}{cov(Z_i, X_i)} = \beta + \frac{cov(Z_i, u_i)}{cov(Z_i, X_i)}$$

- One way to implement instrumental variables estimation is via two-stage least squares.
- In the first stage, charter school attendance is regressed on the instrument. In the second stage, test scores are regressed on the predicted value of charter school attendance.

Correction 3: Panel Data Fixed Effects

- In this case, the researcher has panel data, which track individuals over time. If charter school attendance varies over time, then one can estimate a fixed effects model

$$Y_{it} = \alpha_i + \beta X_{it} + u_{it}$$

Correction 3: Panel Data Fixed Effects

- In this case, the researcher has panel data, which track individuals over time. If charter school attendance varies over time, then one can estimate a fixed effects model

$$Y_{it} = \alpha_i + \beta X_{it} + u_{it}$$

- With only two time periods, fixed effects estimation is equivalent to a regression of changes on changes:

$$\Delta Y_{it} = \beta \Delta X_{it} + \Delta u_{it}$$

Correction 4: Difference in difference

- In this case, one uses a policy change, such as the opening of a new charter school, and creates a treatment and control group. In this case, the regression model can be written as:

$$Y_{it} = \alpha + \beta_1 Treat_i + \beta_2 After_t + \beta_3 Treat_i \times After_t + u_{it}$$

Correction 4: Difference in difference

- In this case, one uses a policy change, such as the opening of a new charter school, and creates a treatment and control group. In this case, the regression model can be written as:

$$Y_{it} = \alpha + \beta_1 Treat_i + \beta_2 After_t + \beta_3 Treat_i \times After_t + u_{it}$$

- The difference in difference estimator can also be calculated in a simple table:

	Treatment	Control	Difference
After	$\alpha + \beta_1 + \beta_2 + \beta_3$	$\alpha + \beta_2$	$\beta_1 + \beta_3$
Before	$\alpha + \beta_1$	α	β_1
Difference	$\beta_2 + \beta_3$	β_2	β_3

Correction 4: Difference in difference

- In this case, one uses a policy change, such as the opening of a new charter school, and creates a treatment and control group. In this case, the regression model can be written as:

$$Y_{it} = \alpha + \beta_1 Treat_i + \beta_2 After_t + \beta_3 Treat_i \times After_t + u_{it}$$

- The difference in difference estimator can also be calculated in a simple table:

	Treatment	Control	Difference
After	$\alpha + \beta_1 + \beta_2 + \beta_3$	$\alpha + \beta_2$	$\beta_1 + \beta_3$
Before	$\alpha + \beta_1$	α	β_1
Difference	$\beta_2 + \beta_3$	β_2	β_3

- Thus, the coefficient on $Treat_i \times After_t$ is the difference in difference estimator.