

実験・準実験 (2)

別所俊一郎

2006年7月14日

Today's attraction

- 説明変数を追加した差推定量
- 差の差 (DiD) 推定量
- テネシー州での少人数教育の実験

実験データを用いた因果効果の回帰分析

- 理想的な無作為割当て実験であれば,

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

を OLS 推定すれば一致推定量を得る

- 有効 (efficient) とは限らない
 - 実際の実験ではしばしば $E[u_i | X_i] \neq 0$ となり, OLS 推定量にバイアス
- 回帰分析を用いた推定
 - 理想的な実験でのより有効な推定量
 - 内的妥当性がないときの不偏・一致推定量
 - 説明変数が追加された差推定量 Difference estimator with additional regressors

説明変数が追加された差推定量

- 実験の結果に影響する他の要因 (W_{1i}, \dots, W_{ri}) が観察可能
 - 学歴 (職業訓練), 年齢・体重・性・既往症 (治験)
- W_{1i}, \dots, W_{ri} は政策によって変化しないとする
- 線形関係を仮定すると

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_{1i} + \dots + \beta_{1+r} W_{ri} + u_i$$

- 重回帰モデルの 4 つの仮定が成り立てば OLS 推定量は一致性
 - とくに $E[u_i | X_i, W_{1i}, \dots, W_{ri}] = 0$
 - この仮定が説得的でないこともしばしば：職業訓練の例
- より弱い仮定で一致性：conditional mean independence
 - $E[u_i | X_i, W_{1i}, \dots, W_{ri}] = \gamma_0 + \gamma_1 W_{1i} + \dots + \gamma_r W_{ri}$
 - 条件つき期待値が W_{1i}, \dots, W_{ri} に依存するが X_i から独立
 - 右辺はより複雑な関数でもかまわない

Conditional Mean Independence

- W_{1i}, \dots, W_{ri} の条件付きで, 政策 X_i が無作為割当て
 - W_{1i}, \dots, W_{ri} で区切られたブロックごとに無作為
 - Block randomization, X_i は conditionally random
- このとき, OLS 推定量 $\hat{\beta}_1$ は一致推定量
 - 政策割当ての確率が W_{1i}, \dots, W_{ri} に依存することを制御済み
 - W_{1i}, \dots, W_{ri} の係数推定量は一般には一貫性を持たない
$$\beta_1 = E[Y_i | X_i = 1, W_{1i}, \dots, W_{ri}] - E[Y_i | X_i = 0, W_{1i}, \dots, W_{ri}]$$
- u_i の平均値は W_{1i}, \dots, W_{ri} を所与とすれば実験群と対照群で同じ
 - W_{1i}, \dots, W_{ri} によって u_i の平均は変化
- W_{1i}, \dots, W_{ri} は政策の結果ではない
 - 結果を表す変数なら内生性を持つ
 - 実験を行う前に決まっていて, 実験に影響されない変数

説明変数が追加された差推定量

1. 有効性 (Efficiency)
 - 無作為割当てであれば単回帰の推定量よりも有効
 - 追加的な説明変数があることで誤差項の分散が小さくなる
2. 無作為割当ての確認
 - 割当てが説明変数と相関していれば, 単回帰による差推定量は一致性を持たず, 確率収束する先は真の値にならない
 - 2種類の OLS 推定量の比較で無作為割当てかどうか確認
3. “Conditional” randomization のための修正
 - ブロックごとに無作為なら W_{1i}, \dots, W_{ri} が割当ての確率を修正するために説明変数に必要
4. 政策の割当てが観測可能な W_{1i}, \dots, W_{ri} 以外の要因にも依存するときには, W_{1i}, \dots, W_{ri} を説明変数として追加するだけでは不十分

Differences-in-differences Estimator

- 差の差推定量 (Differences-in-differences Estimator)

- 実験群の結果の変化の平均と対照群の結果の変化の平均の差

$$\hat{\beta}_1^{DD} = \underbrace{\left(\bar{Y}^{TA} - \bar{Y}^{TA} \right)}_{\text{実験群の変化の平均}} - \underbrace{\left(\bar{Y}^{CA} - \bar{Y}^{CA} \right)}_{\text{対照群の変化の平均}} = \Delta \bar{Y}^T - \Delta \bar{Y}^C$$

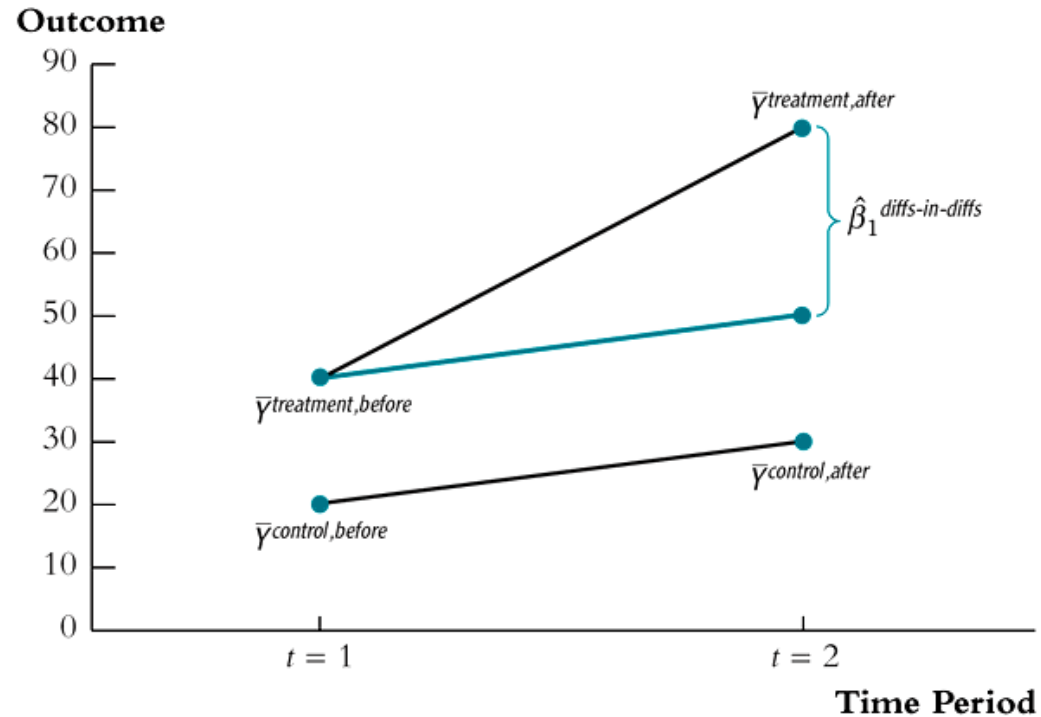
- 割当てが無作為なら $\hat{\beta}_1^{DD}$ は不偏性・一致性を持つ

$$\Delta Y_i = \beta_0 + \beta_1 X_i + u_i$$

- 有効性：無作為割当てなら差推定量より有効．観測不能な通時的に一定な要因があるときはその限りでない
- 実験前の差を除去
 - 実験群と対照群で系統的に異なる初期値への効果を除去
 - 政策が実験前の Y_i と相関し， $E[u_i|X_i] = 0$ のとき，差推定量にバイアス

FIGURE 11.1 The Differences-in-Differences Estimator

The post-treatment difference between the treatment and control groups is $80 - 30 = 50$, but this overstates the treatment effect because before the treatment Y was higher for the treatment than the control group by $40 - 20 = 20$. The differences-in-differences estimator is the difference between the final and initial gaps, so that $\hat{\beta}_1^{diffs-in-diffs} = (80 - 30) - (40 - 20) = 50 - 20 = 30$. Equivalently, the differences-in-differences estimator is the average change for the treatment group minus the average change for the control group, that is, $\hat{\beta}_1^{diffs-in-diffs} = \Delta \bar{Y}^{treatment} - \Delta \bar{Y}^{control} = (80 - 40) - (30 - 20) = 30$.



説明変数が追加された差の差推定量

Differences-in-differences estimator with additional regressors

- 実験前の特性 W_{1i}, \dots, W_{ri} を追加
 - Y_i の変化の差を説明
 - 職業訓練の例：学歴は，実験群と対照群に共通に，賃金上昇率に影響

$$\Delta Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_{1i} + \dots + \beta_{1+r} W_{ri} + u_i$$

- X_i が無作為割り当てなら OLS 推定量 $\hat{\beta}_1$ は不偏
 - X_i が無作為割り当てならより有効な推定量
 - 無作為割り当ての確認が可能
 - Conditional randomization の効果を修正
- 多期間への拡張も可能
 - 固定効果モデルの応用で， W_i と時間固定効果の交差項が必要

異なるグループでの因果効果の推定

- 因果効果の大きさは個人の特性によって異なるかもしれない
 - コレステロール値の下がり方は、すでに低いほうが小さい
 - 職業訓練の効果は女性のほうが大きい / やる気があるほうが大きい
- 個人の特性は観察可能か?
 - 測定可能な特性：交差項 $X_i W_i$ を追加 係数が効果の差
 - 測定不可能な特性：(後述)

部分的な遵守 (partial compliance) への対処

- 政策 X_i と誤差項 u_i が相関 OLS 推定量 $\hat{\beta}_1$ にバイアス
 - やる気のある参加者はプログラムがなくてもよい結果
- 操作変数 Z_i が利用可能なら操作変数法で解決
 - 当初の割当て水準が操作変数となりうる
 - Relevancy : 部分的にでも手続きに従う 実際的水準 X_i と相関
 - Exogeneity : 割り当てが無作為なら , $E[u_i|Z_i] = 0$
 - 差推定量でも差の差推定量でもおなじ

無作為割り当ての検定

- 観測可能な変数に割り当てが依存するか
- Random receipt の検定
 - 無作為割り当てなら, X_i は観測可能な変数と無相関
 - X_i を W_{1i}, \dots, W_{ri} に回帰して係数について仮説検定
 - W_{1i}, \dots, W_{ri} の係数が全てゼロという帰無仮説を F 検定
- Random assignment の検定
 - 無作為割り当てなら, 当初の割り当て水準 Z_i は観測可能な変数と無相関
 - Z_i を W_{1i}, \dots, W_{ri} に回帰して係数について仮説検定
 - W_{1i}, \dots, W_{ri} の係数が全てゼロという帰無仮説を F 検定

少人数教育の効果の実験による推定

- 「少人数学級はテストの成績に効果はあるか?」の実験
 - Project STAR (Student-Teacher Achievement Ratio)
 - 80年代後半, テネシー州の大規模な実験: 4年で1200万ドル
 - 結果の影響力は大きい
- 実験のデザイン
 - 幼稚園から小学3年まで3種類の学級編成を用意
 - * 通常: 1クラス 22~25人
 - * 少人数: 1クラス 13~17人
 - * 補佐付き: 1クラス 22~25人で教師の補佐つき
 - 85~86年に入学した児童を無作為に振り分け
 - 教師も各クラスに無作為に振り分け
 - 毎年, 国語と算数の標準テスト (SAT) を受験
 - 4年で11600人, 80校が参加

少人数教育の効果の実験による推定

実験デザインとの乖離

- 児童は4年間（幼稚園がなければ3年間）同じクラスのはず
- 通常クラスの児童は2年目（小1）で補佐付きへ無作為振分け
 - － 通常クラスの児童の親の反対
 - － 少人数クラスの児童はそのまま
- 10%の児童が相性や行動の問題でクラス替え
 - － ランダムならバイアスの原因とはならない
 - － 教育に熱心な親からの圧力の可能性も
- クラス替え・引越しのため，クラスの児童数が途中で変化

STAR データの分析：差推定量

実験群が 2 つ：少人数クラスと補佐付きクラス

- 差推定量の修正：2 つの「政策ダミー」

$$Y_i = \beta_0 + \beta_1 SC_i + \beta_2 RA_i + u_i$$

- Y_i ：児童 i のテストの点数
- SC_i ：児童 i が少人数クラスにいれば 1，それ以外なら 0
- RA_i ：児童 i が補佐付きクラスにいれば 1，それ以外なら 0
- OLS 推定で β_1, β_2 の推定値を得る (Table 11.1)
 - 幼稚園では，少人数で 13.9 点，補佐付きで 0.31 点上昇
 - 小 1～小 3：少人数クラスの点数上昇効果は統計的に有意
 - 小 2～小 3：補佐付きクラスの点数上昇効果は確認できない

TABLE 11.1 Project STAR: Differences Estimates of Effect on Standardized Test Scores of Class Size Treatment Group

Regressor	Grade			
	K	1	2	3
Small class	13.90** (2.45)	29.78** (2.83)	19.39** (2.71)	15.59** (2.40)
Regular size with aide	0.31 (2.27)	11.96** (2.65)	3.48 (2.54)	-0.29 (2.27)
Intercept	918.04** (1.63)	1,039.39** (1.78)	1,157.81** (1.82)	1,228.51** (1.68)
Number of observations	5,786	6,379	6,049	5,967

The regressions were estimated using the Project STAR Public Access Data Set described in Appendix 11.1. The dependent variable is the student's combined score on the math and reading portions of the Stanford Achievement Test. Standard errors are given in parentheses under the coefficients. **The individual coefficient is statistically significant at the 1% significance level using a two-sided test.

STAR データの分析：説明変数の追加

説明変数を追加するとより有効な推定量を得られる (Table 11.2)

- 「政策」が誤差項と相関していれば，バイアスを除去可能
- 説明変数を追加しても結果は変化せず：無作為割当て
- \bar{R}^2 は大きく，係数の SE は小さく．
- 教師の経験の効果
 - － 教師も無作為に割り当てているから「実験」
 - － 学校内でのみ無作為なので，学校間格差の可能性
 - － 学校ダミーを追加 Conditional mean independence
 - － 10 年の経験で 7.4 点上昇
- 他の係数推定量はバイアスの可能性
 - － 無作為割当てでないので誤差項と相関の可能性
 - － 例：人種・食糧切符は外部教育機会を代理するかも？

TABLE 11.2 Project STAR: Differences Estimates with Additional Regressors for Kindergarten

Regressor	(1)	(2)	(3)	(4)
Small class	13.90** (2.45)	14.00** (2.45)	15.93** (2.24)	15.89** (2.16)
Regular size with aide	0.31 (2.27)	-0.60 (2.25)	1.22 (2.04)	1.79 (1.96)
Teacher's years of experience		1.47** (0.17)	0.74** (0.17)	0.66** (0.17)
Boy				-12.09** (1.67)
Free lunch eligible				-34.70** (1.99)
Black				-25.43** (3.50)
Race other than black or white				-8.50 (12.52)
Intercept	918.04** (1.63)	904.72** (2.22)		
School indicator variables?	no	no	yes	yes
\bar{R}^2	0.01	0.02	0.22	0.28
Number of observations	5,786	5,766	5,766	5,748

The regressions were estimated using the Project STAR Public Access Data Set described in Appendix 11.1. The dependent variable is the combined test score on the math and reading portions of the Stanford Achievement Test. The number of observations differ in the different regressions because of some missing data. Standard errors are given in parentheses under coefficients. The individual coefficient is statistically significant at the *5% level or **1% significance level using a two-sided test.

STAR データの分析：結果の解釈

点数の標準偏差との比較：学年により平均点の差 (Table 11.3)

- 幼稚園：係数 13.9(2.45)，点数の SD は 73.7 $13.9/73.7 = 0.19$
SE は $2.45/73.7 = 0.03$
- 少人数の効果はどの学年でも同じくらい：点数の SD の約 20%
- 補佐付きの効果は，幼稚園・小 2・小 3 でほぼゼロ
- 少人数と補佐付きの差はどの学年でも点数の SD の約 20%
- 小 1 の結果は，たまたま対照群の成績が悪かった？

他の係数との比較

- 男女の点数差 12 点よりは大きい
- 教師の教歴と比べると 20 年分に相当
- かなり大きいと思われる。

TABLE 11.3 Estimated Class Size Effects in Units of Standard Deviations of the Test Score Across Students

Treatment Group	Grade			
	K	1	2	3
Small class	0.19** (0.03)	0.33** (0.03)	0.23** (0.03)	0.21** (0.03)
Regular size with aide	0.00 (0.03)	0.13** (0.03)	0.04 (0.03)	0.00 (0.03)
Sample standard deviation of test scores (s_Y)	73.70	91.30	84.10	73.30

The estimates and standard errors in the first two rows are the estimated effects in Table 11.1, divided by the sample standard deviation of the Stanford Achievement Test for that grade (the final row in this table), computed using data on the students in the experiment. Standard errors are given in parentheses under coefficients. **The individual coefficient is statistically significant at the 1% significance level using a two-sided test.

STAR データの追加的な結果

- 少人数クラスの効果は低学年に集中
 - 小1～小3で、通常クラスと少人数クラスの点数格差は同じ
 - 当初の割当ての効果が持続
- 補佐付き学級の効果はほとんどない
 - 実験手続きに従わない児童の存在によるバイアス？
 - 当初の割当てを IV とする TSLS の結果は OLS と大差ない
 - 実験手続きに従わないことからのバイアスは小さい

観測データと実験データの比較

カリフォルニア・マサチューセッツデータとの比較 (Table 11.4)

- 比較可能とするために, 実験に合わせて 7.5 人減を考察
- CA : 係数 $0.73 \times$ 人数 $7.5 = 5.5$ 点 SD の 0.14 倍 ($=5.5/38$)
- データの SD が違うので信頼区間で比較 : 似ている!

結果の違いの原因

- 観測データの分析に付きまとう内的妥当性への疑義 (バイアス)
 - 観測誤差 : 地区データを利用, 引越し, 越境通学
- 外的妥当性?
 - 観測データでのクラスの人数は実験データと異なる : 越境通学
 - 観測データは 98 年の CA, MA : STAR は 80 年代の南部
 - 観測データは小 4 (MA) と小 5 (CA) : STAR は幼稚園 ~ 小 3
- 結果が似ているほうが驚き : 外的妥当性はあるかも

TABLE 11.4 Estimated Effects of Reducing the Student-Teacher Ratio by 7.5 Based on the STAR Data and the California and Massachusetts Observational Data

Study	$\hat{\beta}_1$	Change in Student-Teacher Ratio	Standard Deviation of Test Scores Across Students	Estimated Effect	95% Confidence Interval
STAR (grade K)	-13.90** (2.45)	Small class vs. regular class	73.8	0.19** (0.03)	(0.13, 0.25)
California	-0.73** (0.26)	-7.5	38.0	0.14** (0.05)	(0.04, 0.24)
Massachusetts	-0.64* (0.27)	-7.5	39.0	0.12* (0.05)	(0.02, 0.22)

The estimated coefficient $\hat{\beta}_1$ for the STAR study is taken from column (1) of Table 11.2. The estimated coefficients for the California and Massachusetts studies are taken from the first column of Table 7.3. The estimated effect is the effect of being in a small class versus a regular class (for STAR) or the effect of reducing the student-teacher ratio by 7.5 (for the California and Massachusetts studies). The 95% confidence interval for the reduction in the student-teacher ratio is this estimated effect ± 1.96 standard errors. Standard errors are given in parentheses under estimated effects. The estimated effects are statistically significantly different from zero at the *5% level or **1% significance level using a two-sided test.