労働経済学

Lecture 10 実証研究における因果的効果の識別
回帰不連続デザイン (Regression Discontinuity Design)

張　俊超

15th June 2017

IV のまとめ

- OLS における説明変数と誤差項との相関を考慮するため、一段階で得られた予測値を二段階の説明変数とし、二段階推定を行うこと。
- IV は一段階のみにあり、二段階から除外される。
- IV は外生的政策変動、災害などの偶発的なものが望ましく、いわゆる自然実験がほしい。(技術上、どんな変数でも二段階推定が可能が、除外制約が満たされない可能性が高い。)
- IV は二段階の説明変数に相関がなかればならないが、その誤差項との相関があると良い IV でなくなる。(親の教育は子供の教育と相関するが、子供の能力とも相関がある。IV として使えない。)

RDD とは

何か観察できる変数がある閾値を超えた時に処置が割り当てられる場合に、その閾値の前後での差を見ることで処置効果を推計する方法。この推定方法は Regression Discontinuity Design（回帰不連続）という。処置の割り当てを決める変数を assignment variable/forcing variable/running variable という。

二種類の RD がある。（回帰不連続は RD または RDD と略す）

- Sharp RD: assignment 変数が閾値超えると必ず処置
- Fuzzy RD: assignment 変数が閾値超えると確率的に処置

Sharp RD

- **閾値を超えると必ず処置を受ける**
  例：100 **点満点の試験の点数が** 95 **点以上の人に必ず奨学金を出す。**

$$D_i = \begin{cases} 1 & \mathit{if} \quad S_i \geq 95 \\ 0 & \mathit{if} \quad S_i < 95 \end{cases}$$

- D_i: **奨学金を受けるかどうか。処置変数。**
- S_i: **点数。**assignment **変数。**
- **点数がわかれば、必ず奨学金を受けるかどうかがわかる。**

Sharp RD

奨学金が大学進学率 y_i に与える影響を見るために、assignment 変数の性質によって、

$$E(Y_{0i}|S_i) = \alpha + \beta S_i$$
$$Y_{1i} = Y_{0i} + \pi$$

回帰分析で表現すれうば、

$$Y_i = \alpha + \beta S_i + \pi D_i + \varepsilon_i$$

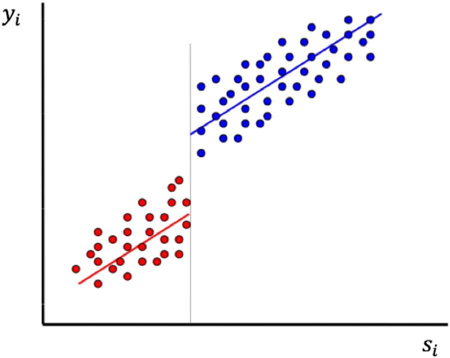
D_i は S_i と相関するだけではなく、s_i に関する関数である（s_i の値によって D_i が決まる）。

π は RD の推定量。$\bar{S}$ は 95 点。

$$\pi = E(Y|S = \bar{S}) - lim_{s \to \bar{S}} E(Y|S = s)$$

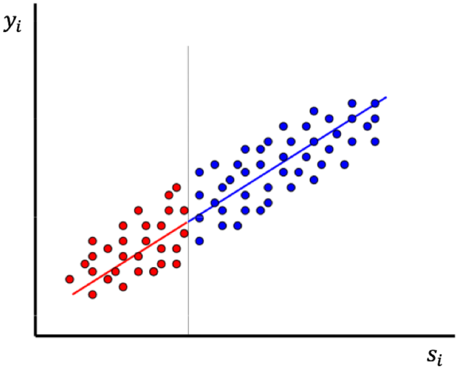
図で見る Sharp RD

処置効果がある場合、s_i の cutoff に不連続が出る。



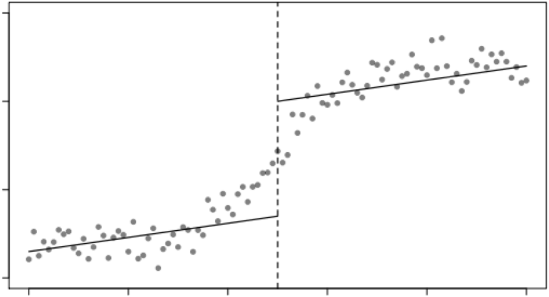
図で見る Sharp RD

処置効果がない場合、s_i の cutoff に不連続がない。



非線形の Sharp RD

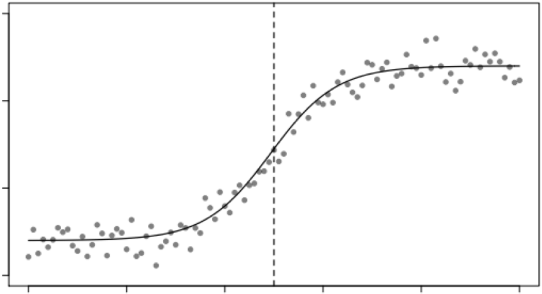
結果変数 Y_i と assignment 変数との間に非線形の相関を持つが、線形 RD で推定してしまうと



推定結果はけっこうバイアスがかかる。RD モデルを柔軟に設定し、非線形モデルを推定したほうがいい。

非線形の Sharp RD

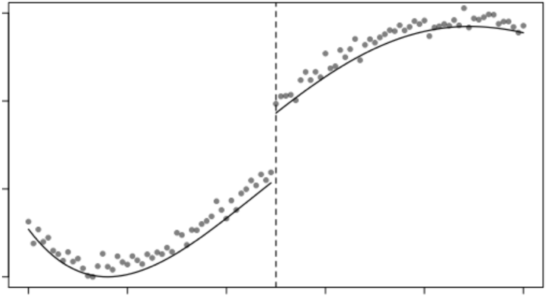
非線形モデル推定により、不連続は見れず、処置効果がないかもしれない。



非線形の Sharp RD

非線形 RD を推定するために、S_i の代わりに、S_i の二乗まで回帰式に投入する（結果変数と assignment 変数は二次関数の関係であれば）

$$Y_i = \alpha + \beta S_i + \gamma S_i^2 + \pi D_i + \varepsilon_i$$



非線形の Sharp RD

しかし、通常、Y_i と S_i との関係は知らない。もっともらしい RD 推定を行うために、$f(S_i)$ は二つの方法でフィットできる。

$$Y_i = \alpha + \pi D_i + f(S_i) + \varepsilon_i$$

- pth Order Polynomial(**多項回帰式？**)
- Local Linear Regression

pth Order Polynomial

pth Order Polynomial による RD 推定は基本的に、

$$Y_i = \alpha + \beta_1 S_i + \beta_2 S_i^2 + ... + \beta_p S_i^p + \pi D_i + \varepsilon_i$$

cutoff の両辺に、異なる関数の形まで緩めることもできる。その場合、cutoff の左辺は

$$E(Y_{0i}) = \alpha + \beta_{01}\tilde{S}_i + \beta_{02}\tilde{S}_i^2 + ... + \beta_{0p}\tilde{S}_i^p$$

cutoff の右辺は

$$E(Y_{1i}) = \alpha + \pi + \beta_{11}\tilde{S}_i + \beta_{12}\tilde{S}_i^2 + ... + \beta_{1p}\tilde{S}_i^p$$

ただし、$\tilde{S}_i = S_i - \bar{S}$、つまり、$S_i$ を $\bar{S}$ について中心化 (centering) すること。

pth Order Polynomial

cutoff 両端の関数の形が異なってもいい RD 推定は、

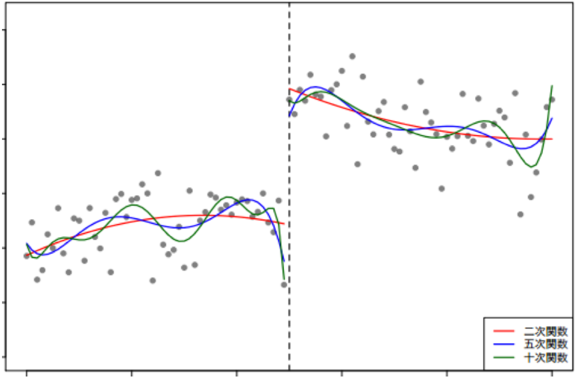
$$
\begin{aligned}
Y_i = & \alpha + \beta_{01}\tilde{S}_i + \beta_{02}\tilde{S}_i^2 + \dots + \beta_{0p}\tilde{S}_i^p + \pi D_i + \\
& + \gamma_1 D_i\tilde{S}_i + \gamma_2 D_i\tilde{S}_i^2 + \dots + \gamma_p D_i\tilde{S}_i^p + \varepsilon_i
\end{aligned}
$$

つまり、追加的に、処置変数と assignment 変数の交差項を入れる。

$\tilde{S}_i$ の中心化により、

π は常に推定したい RDD の推定量。

pth Order Polynomial

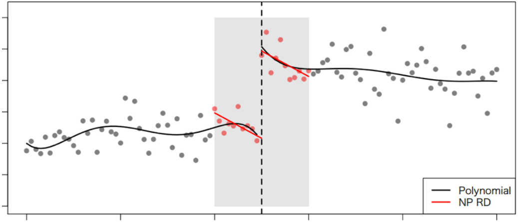


Local Linear Regression

Polynomial 法は多項式により、円滑な関数で $f(S_i)$ をフィットする。その他、ノンパラの Local Linear Regression でも $f(S_i)$ をフィットする方法がある。Local Linear Regression を推定するために、

$$Y_i = \alpha + \pi D_i + \beta \tilde{S}_i + \gamma \tilde{S}_i D_i + \varepsilon_i$$

つまり、通常のモデルに、S_i と D_i との交差項を入れる。（二乗、三乗...はいない）ただし、cutoff 近傍のサンプルを扱うこと。



Local Linear Regression

近傍のバンド幅 (bandwidth) **の選択は問題になるが、一般的に、いくつかのバンド幅で推定して、その推定値の頑健性をチェックすることができる。**

Lee and Lemieux (2010) **は二つのバンド幅を選択する方法を提案した。興味のある方、英語論文を読んでください。ここで略す。**

RD に必要な仮定

1. Discontinuous rule がある、奨学金の例では、95 点以上と 95 点未満の間で奨学金の支給率がはっきり不連続になっている。
2. Assignment variable, つまりテストの点数がはっきりわかる。
3. Positive density of assignment variable at the threshold つまり、$f_s(\bar{S}) > 0$ が 95 点付近の人がちゃんといる
4. Continuous impact of assignment variable on outcome つまり、試験の点数が上がるにつれて大学進学率が (奨学金と関係なく) 上がって行ってもいいけど、その変化は連続的。
5. Continuous density; incomplete control of assignment variable つまり、試験の点数の分布が threshold の前後で偏ったりしていない = 完全にコントロールできない = ランダンム

RD のポイント

- Assignment variable が個人の努力によって多少上げたり下げたりできるとしても、完全にはコントロールできない場合は、threshold の近傍ではランダムに分布しているとみなすことができる。つまり、ごく狭い範囲では random assignment とみなすことができる。
- したがって、assignment variable が誤差なく完全に操作可能である場合は、RD は使えない。
  - テストの点数は、95 点を狙っていても 94 点しか取れない場合もあるし、92 点くらいの実力の子がまぐれで 95 点以上とることもあるので OK.
  - 不確定要素が全くないものだけがアウト

Fuzzy RD

Assignment 変数が閾値を超えると、確率的に処置を受ける。

$$P(D_i = 1|S_i) = \begin{cases} g_1(S_i) & \textit{if} \quad S_i \geq 95 \\ g_0(S_i) & \textit{if} \quad S_i < 95 \end{cases} \textit{where} \quad g_1(S_i) \neq g_0(S_i)$$

奨学金のケースでは、95 点以上でも必ず奨学金をもらえることはない。95 未満でももらえないことはない。ただし、95 以上と 95 未満のグループ間、もらえる確率がはっきり異なる（不連続がある)。(95 以上であれば 70% でもらえる、95 未満は 5% でもらえる。)

よって、その不連続を IV(操作変数) として扱い、二段階推定で奨学金が大学進学率に与える効果を推定できる。

処置の確率と点数の関係が、$P(D_i = 1|S_i) = g_0(S_i) + [g_1(S_i) - g_0(S_i)]T_i$, $T_i = 1(S_i > \bar{S})$

Fuzzy RD=IV

$$Y_i = \alpha + \beta S_i + \pi D_i + \varepsilon_i \tag{1}$$

$$D_i = \delta + \gamma S_i + \tau T_i + \epsilon_i \tag{2}$$

- (2) は一段階。
- T_i(95 点以上のダミー変数) が IV として用いて、奨学金を受けるかどうかを予測する。
- D_i の予測値を得て、(1) 式に代入し、大学進学率に与える効果が分析できる。

Fuzzy RD の仮定

Fuzzy RD は IV と一緒などで、IV の仮定を満たさないといけない。つまり、

1. $Cov(D_i, T_i) \neq 0$
   $\tau \neq 0$ かつ統計的に有意であれば、OK

2. $Cov(T_i, \varepsilon_i) = 0$
   除外制約

Fuzzy RD の拡張版

Sharp RD は Polynomial 法と Local Linear Regression の拡張的なモデルで推定できる。Fuzzy RD も同様、同じような設定で推定できる。以下の例は Local Linear Regression による Fuzzy RD 推定。（Polynomial による例は略す）

$$Y_i = \alpha + \pi D_i + \beta \tilde{S}_i + \varepsilon_i$$

$$D_i = \delta + \gamma T_i + \tau \tilde{S}_i + \theta \tilde{S}_i T_i + \epsilon_i$$

T_i と $S_i T_i$ は除外される IV になる。

Angrist and Lavy(1999)

Angrist and Lavy(1999) は Fuzzy RD で、学級規模がテストの点数に与える因果効果を推定した。

その研究では、

1. 一段階では、平均的な学級規模（連続変数）の不連続を用いた。
2. 複数の不連続を用いた。

一般的に、学級規模が小さければ小さいほど、学生たちのテスト点数が高いと予測される。他の要因を一定であれば、小さい学級規模では、教師がもっと多くの学生に細かく指導できる。学生一人当たりに配分された指導の時間が長くなる。

Angrist and Lavy(1999)

OLS で推定すると、学級規模が観測されない要因と相関し、推定される学級規模の効果がバイアスがかかる。Angrist and Lavy(1999) は Talmudic rule による学級規模の不連続を用いて推定した。

イスラエルでは、学級規模の上限は 40 人。

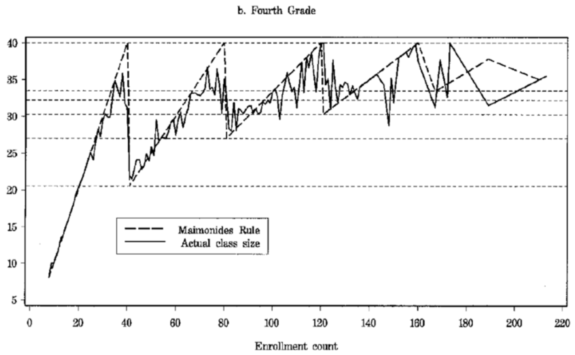
- ある学校に 40 人の学生がいる。クラスは一つだけ、学級規模が 40 人。
- ある学校に 41 人の学生がいる。クラスは二つ、それぞれ学級規模は 21 人と 20 人。

登録学生数がわかれば、平均的な学級規模が計算できる。学校はそのルールに従い、クラスを分けるべきである。

$$m_{sc} = \frac{e_s}{int[\frac{(e_s-1)}{40}] + 1}$$

Angrist and Lavy(1999)

しかし、何らかの事情で、そのルールを従わない学校がある。（assignment 変数の値により、完全に処置を決めない。）Sharp RD は使えない。Angrist and Lavy(1999) は規定された学級規模を実際の学級規模の IV として用いて、Fuzzy RD 推定を行った。



Angrist and Lavy(1999)

$$Y_{isc} = \alpha_0 + \alpha_1 d_s + \rho n_{sc} + \beta_1 e_s + \beta_2 e_s^2 + \eta_{isc}$$
$$n_{sc} = \gamma_0 + \gamma_1 d_s + \pi m_{sc} + \delta_1 e_s + \delta_2 e_s^2 + \xi_{isc}$$

- Y_{isc} は s 学校、c クラスの学生 i の点数。
- n_{sc} は実際の s 学校の c クラスの学級規模。関心を持つ変数。
- m_{sc} は学校 s、クラス c の規定された学級規模。（c の小文字は付いてるけど、実は学校内では、規定された学級規模に Variation はない。）
- e_s は学校 s の登録学生数。
- n_{sc} は D_i、e_s は X_i、m_{sc} は T_i に対応する。（Fuzzy RD のページに参照）

Angrist and Lavy(1999) 推定結果

TABLE IV
2SLS ESTIMATES FOR 1991 (FIFTH GRADERS)

	Reading comprehension						Math					
	Full sample				+/− 5 Discontinuity sample		Full sample				+/− 5 Discontinuity sample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mean score		74.4			74.5			67.3			67.0	
(s.d.)		(7.7)			(8.2)			(9.6)			(10.2)	
Regressors												
Class size	−.158	−.275	−.260	−.186	−.410	−.582	−.013	−.230	−.261	−.202	−.185	−.443
	(.040)	(.066)	(.081)	(.104)	(.113)	(.181)	(.056)	(.092)	(.113)	(.131)	(.151)	(.236)
Percent disadvantaged	−.372	−.369	−.369		−.477	−.461	−.355	−.350	−.350		−.459	−.435
	(.014)	(.014)	(.013)		(.037)	(.037)	(.019)	(.019)	(.019)		(.049)	(.049)
Enrollment		.022	.012			.053		.041	.062			.079
		(.009)	(.026)			(.028)		(.012)	(.037)			(.036)
Enrollment squared/100			.005						−.010			
			(.011)						(.016)			
Piecewise linear trend				.136						.193		
				(.032)						(.040)		
Root MSE	6.15	6.23	6.22	7.71	6.79	7.15	8.34	8.40	8.42	9.49	8.79	9.10
N		2019		1961	471			2018		1960	471	

The unit of observation is the average score in the class. Standard errors are reported in parentheses. Standard errors were corrected for within-school correlation between classes. All estimates use f_{sc} as an instrument for class size.

Angrist and Lavy(1999) 推定結果

TABLE V
2SLS ESTIMATES FOR 1991 (FOURTH GRADERS)

	Reading comprehension						Math					
	Full sample				+/− 5 Discontinuity sample		Full sample				+/− 5 Discontinuity sample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mean score		72.5			72.5			67.3			68.7	
(s.d.)		(8.0)			(7.8)			(9.6)			(9.1)	
Regressors												
Class size	−.110	−.133	−.074	−.147	−.098	−.150	.049	−.050	−.033	−.098	.095	.023
	(.040)	(.059)	(.067)	(.084)	(.090)	(.128)	(.048)	(.070)	(.081)	(.092)	(.114)	(.160)
Percent disadvantaged	−.346	−.345	−.346		−.354	−.347	−.290	−.284	−.284		−.299	−.290
	(.014)	(.014)	(.014)		(.034)	(.034)	(.017)	(.017)	(.017)		(.042)	(.043)
Enrollment		.005	−.040			.017		−.020	.007			.023
		(.008)	(.024)			(.022)		(.010)	(.029)			(.028)
Enrollment squared/100			.021						.006			
			(.011)						(.014)			
Piecewise linear trend				.100						.130		
				(.026)						(.028)		
Root MSE	6.65	6.66	6.63	8.02	6.64	6.69	7.82	7.82	7.82	8.65	8.23	8.24
N		2049		2001	415			2049		2001	415	

The unit of observation is the average score in the class. Standard errors are reported in parentheses. Standard errors were corrected for within-school correlation between classes. All 2SLS estimates use f_{sc} as an instrument for class size.

Angrist and Lavy(1999),OLS との比較

TABLE II
OLS ESTIMATES FOR 1991

	5th Grade						4th Grade					
	Reading comprehension			Math			Reading comprehension			Math		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mean score		74.3			67.3			72.5			69.9	
(s.d.)		(8.1)			(9.9)			(8.0)			(8.8)	
Regressors												
Class size	.221	−.031	−.025	.322	.076	.019	0.141	−.053	−.040	.221	.055	.009
	(.031)	(.026)	(.031)	(.039)	(.036)	(.044)	(.033)	(.028)	(.033)	(.036)	(.033)	(.039)
Percent disadvantaged		−.350	−.351		−.340	−.332		−.339	−.341		−.289	−.281
		(.012)	(.013)		(.018)	(.018)		(.013)	(.014)		(.016)	(.016)
Enrollment			−.002			.017			−.004			.014
			(.006)			(.009)			(.007)			(.008)
Root MSE	7.54	6.10	6.10	9.36	8.32	8.30	7.94	6.65	6.65	8.66	7.82	7.81
R^2	.036	.369	.369	.048	.249	.252	.013	.309	.309	.025	.204	.207
N		2,019			2,018			2,049			2,049	

The unit of observation is the average score in the class. Standard errors are reported in parentheses. Standard errors were corrected for within-school correlation between classes.

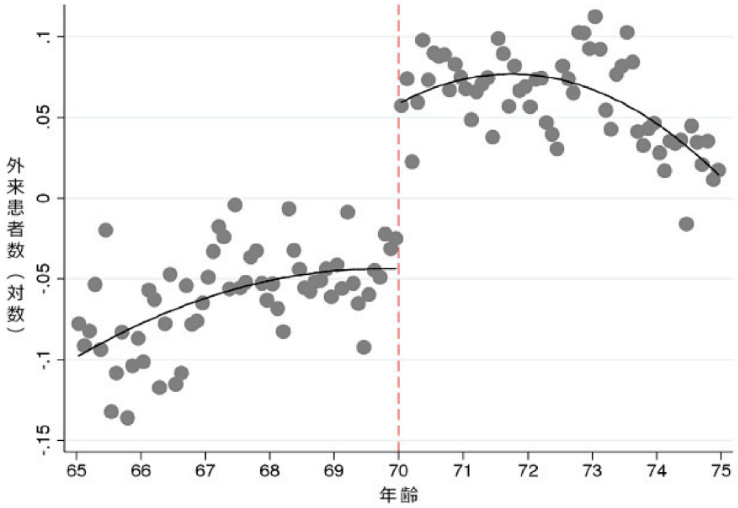
Angrist and Lavy(1999) のまとめ

1. OLS 単回帰で学級規模がテスト点数に与える効果を推定すると、正の効果を得る。学級規模が大きいほうがいい？見かけ上の相関。
2. Fuzzy RD(IV) で推定すると、負の効果を得る。学級規模が小さいければ小さいほど、学生のテスト点数が高い。
3. OLS と Fuzzy RD の違いによる、学級規模の内生性を確認した。(お金持ちの家族は子供を小さい学級規模を持つ学校に入学させている？)

Shigeoka(2014)：日本の RD

- Shigeoka(2014) は Sharp RD で「高齢者 1 割負担」が医療施設の利用、個人の健康などに与える効果を分析した。
- アメリカでは、社会実験でランダムに負担割合を割り振ることで、処置効果を分析した文献がある。
- Shigeoka(2014) は日本の「70 歳自己負担割合が変わる」という制度に着目し、「自然実験」として利用し、Sharp RD でその処置効果を推定した。
- 69 歳の負担割合が 3 割、70 歳の負担割合が 1 割。

Shigeoka(2014)：日本の RD



Shigeoka(2014)：日本の RD

- 医療施設の利用が劇的に増加したが、健康的結果変数にほとんど有意な結果は出ていない。
- 政府の予算が一定であれば、医療支出を増やすと、必ず他の予算を減らしないといけない。
- 「1 割負担」による医療支出の増加について、年寄りの健康は改善されていない。政策は間違った？