

# Chapter4

## Regression Discontinuity Designs

1

### 目的

- ・人間の行動は様々な規則によって縛られている。

例：カリフォルニア州の一クラスあたりの生徒数

退職手当の受給許可開始年齢、アメリカ軍の入隊許可など

- ・これらのルールが興味深い実験を生み出す。

・これらはRDDによって解析され、（局所的に）ほとんどランダム処置と同じような因果効果を導く

2

## 4.1 誕生日と葬式と

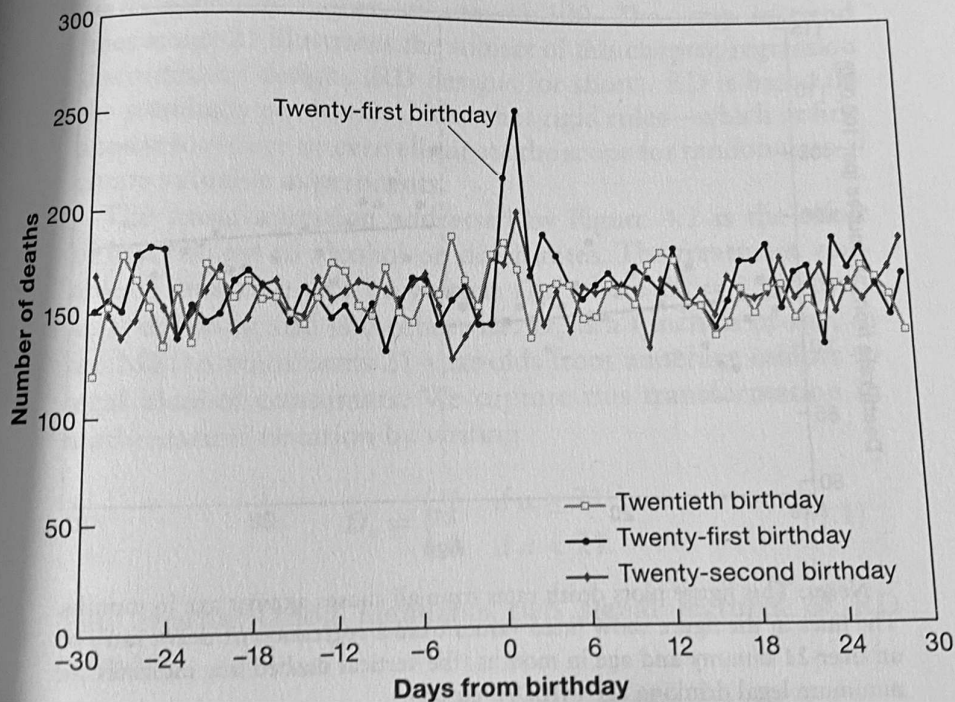
- ・アメリカでは21歳で飲酒可能となる。
- ・一方で、アメリカの大学長らによるMLDA (Minimum legal drinking age) の18歳 (ベトナム戦争時代の水準) への引き下げのためのロビー活動が行われている。
- ・背景理論はthe Amethyst Initiative として知られ、内容は18歳での飲酒が暴飲を防ぎ、成熟した飲酒の文化を促すというもの。

3

- ・MLDAの変化の歴史が二つの自然実験を提供している。  
(二つ目の自然実験は第五章で紹介済み)
- ・一つ目の自然実験では年齢のわずかな変化が法的に認められた飲酒開始に多大に影響することを示している。

4

FIGURE 4.1  
Birthdays and funerals



右図は横軸に誕生日からの日数。

(0が誕生日当日、+5だと誕生日から五日後の意味)

縦軸は死者の数を表す。

各折れ線はそれぞれ20、21、22歳の誕生日の周辺を表している。

(データは1997年から2003年のもの)

5

・図に見るように21歳の誕生日後、数日にかけて死亡者数が急上昇していることがわかる。(約100ほどの上昇)

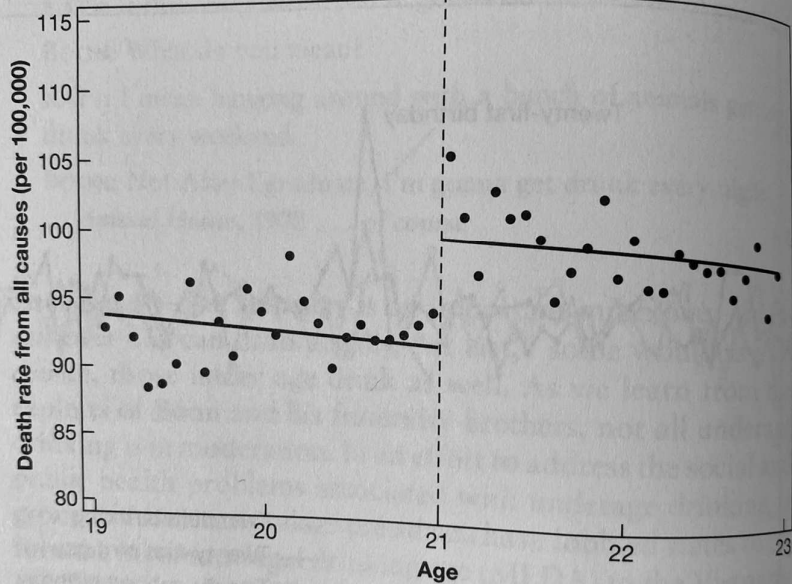
・この効果は20歳、および22歳の誕生日後のトレンドを見るに誕生日会での飲酒だけによるものではないと推察される。

・しかし、この21歳での死亡者の増加はMLDAによるものなのか、また、この効果は長く続くような重大な効果なのか、まだ考察の余地がある。

6

# Sharp RD

FIGURE 4.2  
A sharp RD estimate of MLDA mortality effects



この図は21歳の誕生日を中心として一か月ごとの死亡率をとったもの。

21歳の地点を境に急かつ持続する死亡率の上昇がみられる。

21歳の左側では95を超えるデータが少ないのに対して、21歳の右側では95を下回るものが少なくなっている。

7

・死亡率はやや下降トレンドにあり、このトレンドを利用して境界の左側から21歳での死亡率を推定するとだいたい92あたりになることが予想される。

・しかしこの値は飲酒が許可されていない年齢からの推定であり、第一章の言葉を使えば観察されていない反事実の結果 $Y_{i0}$ とみなせる。

・実際は、100に近い値をとり、この二つの値の違いがこの章での主題であるRDDを明示している。

・RD（不連続回帰）は厳しい規制というものが価値ある実験を作り出すという一見すると矛盾している発想をもとにしている。

8

# 定式化

- ・この時、処置効果を以下のように定義する

$$D_a = \begin{cases} 1 & \text{if } a \geq 21 \\ 0 & \text{if } a < 21 \end{cases}$$

つまり $D_a = 1$ はその人が飲酒可能年齢に達していることを表す。

- ・RDの二つの特徴

- 1、処置効果は $a$ によってのみ決まり、 $a$ が既知であれば $D_a$ が決定される。
- 2、 $D_a$ は $a$ による不連続な関数となる。

- ・処置変数を決定するような変数をrunning variableと呼ぶ。  
(今回の例では年齢)

9

・Sharp RDにおいてはrunning variableが境を超すことで明確に処置効果が切り替わる。

・MLDAの例では、年齢が21歳になると処置変数が0から1に切り替わるのでSharp RDの研究対象となる。

・後半では確率などによって不連続が生まれるモデルであるFuzzy RDについて考察する。

10

・死亡率は明らかにMLDAとは関連していない原因によって年齢とともに変化する。

例えば、がんなどの内的原因。交通事故、自殺などの外的原因

・こういったトレンドの変化をMLDAによる潜在的な効果と分けるために、

RDにおける解析で年齢によって生じたトレンドを制御する。

11

## 定式化

・回帰式は以下のようなになる。

$$\bar{M}_a = \alpha + \rho D_a + \gamma a + e_a$$

$\bar{M}_a$ :月aにおける死亡率、 $D_a$ :処置変数、a:年齢（月単位）

・ $\gamma$ は年齢による死亡率のトレンドを表す。

・ $\rho$ はMLDAの境界における上がり幅の推定値（因果効果）

・この式における $\rho$ の推定値は7.7

・平均的な死亡率が9.5（10万人あたり）であることを考えるとこの因果効果は大きい。

12

# 欠落変数バイアス

- ・  $D_a$  と新たに含められた説明変数との共分散に依存して推定結果  $\rho$  は変化する。（欠落変数バイアス）

- ・ しかしながら、 $D_a$  は年齢のみに依存し、年齢は回帰式内において制御されているので、年齢の死亡率における効果が線形式によってあらわされるのであれば、欠落変数問題は生じえないだろう。

13

- ・ 今回での議論の中心は、割当変数（running variable）と結果の関係（因果効果）が線形式で制御された年齢を含む回帰式によって正しく推定されるかどうか。

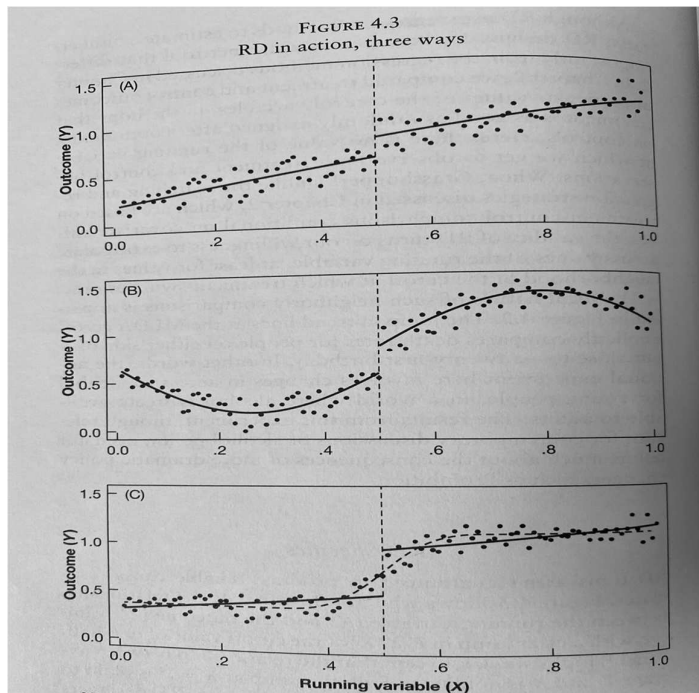
- ・ RDは回帰式を用いているが一般の回帰式とは違い有用な因果推定のための道具としてみなされている。

（Chapter2におけるマッチングなど）

- ・ ただし、この因果効果は境界付近における局所的な効果のみ伝える。よってここでの議論はMLDAにおける飲酒の危険性のみの評価である。

14

# 不連続回帰の特定



・常にRDの方法で信頼できる因果効果を推定できるわけではない。

・それは左図のように因果効果がなくても明確なジャンプがみられるような場合である。

・Cの例は特に厄介（不連続でないのに不連続であると認識してしまう。）

15

## RDの失敗に対する二つの対策

・一つ目のモデル：  
非線形性をそのまま制御する、つまり回帰式に含める。

・二つ目のモデル：  
境界近くのデータにのみ着目する。

16



# 一つ目のモデル

・非線形モデルは通常、running variableの多項式を回帰式に含めることによって制御する。

・どれだけ高次の多項式まで含めるかどうかは考察が必要。

・モデルが研究者自身の偏見をそのまま表現してる場合には注意が必要となる。

・二次の年齢の項を含めた回帰式が以下である。

$$\bar{M}_a = \alpha + \rho D_a + \gamma a + \gamma a^2 + e_a$$

17

境界面の左側・右側に関して割り当て変数にかかわる係数を異なるように設定するモデルを考えることができる。

例えば以下のようなもの:

$$\bar{M}_a = \alpha + \rho D_a + \gamma(a - a_0) + \delta[(a - a_0)D_a] + e_a$$

この場合、年齢の変数 $a$ を $(a - a_0)$ に置き換えた。

$a_0$ は境界でのrunning variableの値を意味する。

今回は $a_0 = 21$ となる。

このように境界の右側・左側で年齢による死亡率のトレンドが変化すると考えるのは、飲酒によって何らかの死亡率のトレンドの変化が存在すると考えられる、あるいはその余地を残すため。

例えば、飲酒が死亡率の下降トレンドを弱めるかもしれない

18

このinteraction Termを含めることによって得られる含意は境界から離れたところにおけるMLDAによる因果効果が

$$\rho + \delta(a - a_0)$$

によって与えられることである。

これは次のように求められる

$$[\alpha + \rho + (\gamma + \delta)(a - a_0)] - [\alpha + \gamma(a - a_0)] = \rho + \delta(a - a_0)$$

左式の第一項は境界の右側において $D_a = 1$ だった時の推定値、第二項は $D_a = 0$ だった時の推定値。

19

・この推定は大胆なものではあるが、21歳よりもある程度離れた年齢にMLDAが設定されている例がないので、この推定を裏付けするよなEvidenceは得られない。

・一方で21歳付近では、飲酒することができる、あるいはできないといった違い以外には大きな差がないように思える。したがって、因果効果は $\rho$ とするのが望ましい。

20

先に紹介したモデルと組み合わせると以下のようになる。

$$\begin{aligned}\bar{M}_a = & \alpha + \rho D_a + \gamma_1(a - a_0) + \gamma_2(a - a_0)^2 \\ & + \delta_1[(a - a_0)D_a] + \delta_2[(a - a_0)^2 D_a] + e_a\end{aligned}$$

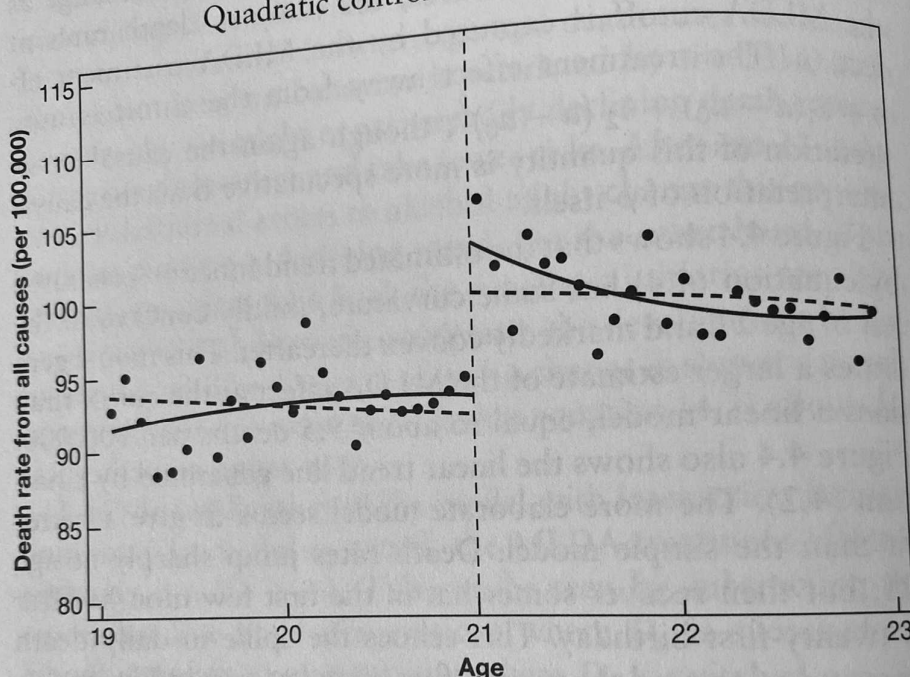
境界付近での因果効果は $\rho$ 、

境界から遠く離れた部分では $\rho + \delta(a - a_0) + \delta_2(a - a_0)^2$ となる。

ただしこの因果効果は精度が低い

21

FIGURE 4.4  
Quadratic control in an RD design



前のスライドのモデルの推定結果。

このモデルは線形トレンドの時よりも因果効果の推定値として大きな値を返す。

このモデルは単純なモデルよりもよりデータを反映している。

一方で技術的なコストがかかる。

22

- ・どちらのモデルがよいのかという問いに対して一般的な規則はないし、データを考察する代用手段もない。

- ・モデリングによって結果があまり変わらないのであれば、(推定結果の頑健性が保たれることになるので) よい。

- ・今回のケースではあまり変わらない。

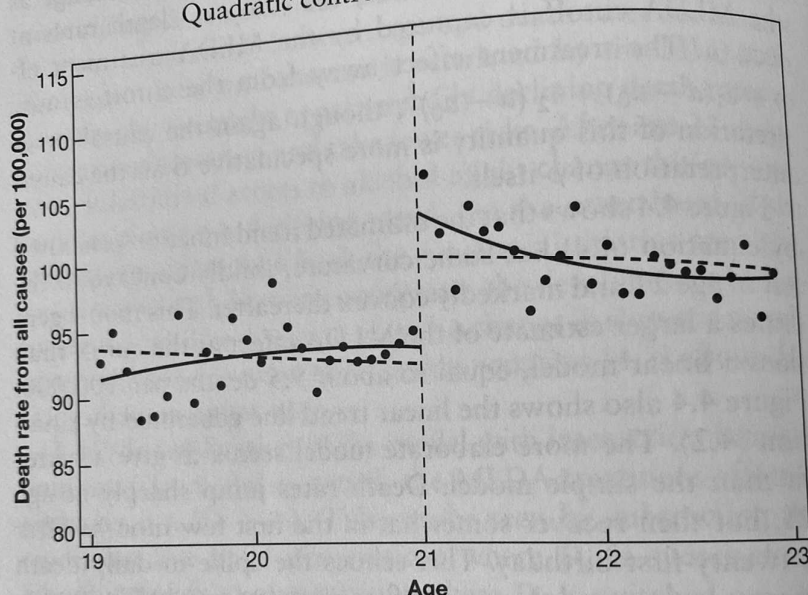
単純なケースでは、境界面で正しく因果効果を推定するのに柔軟に見える。

複雑なケースでは、境界面のすぐ右側における死亡率の変化を正確にとらえてる。

23

- ・境界面での因果効果しか推定できないが、飲酒開始が死亡率に与える影響を考察することは可能である。

FIGURE 4.4  
Quadratic control in an RD design



- ・右図からもわかるように境界面の右側では左側に比べ死亡率が概して高い。

- ・このことは飲酒開始が、死亡率に対する持続的な効果を持ち得ることを示唆する。

24

# ジャンプがどの程度飲酒によるものなのか？

- ・ 死亡原因別にみて飲酒と死亡原因の関連から飲酒のジャンプに対する影響を鑑みる。
- ・ ジャンプがみられないような死亡原因があれば、因果関係がより明確になる。

## 結果

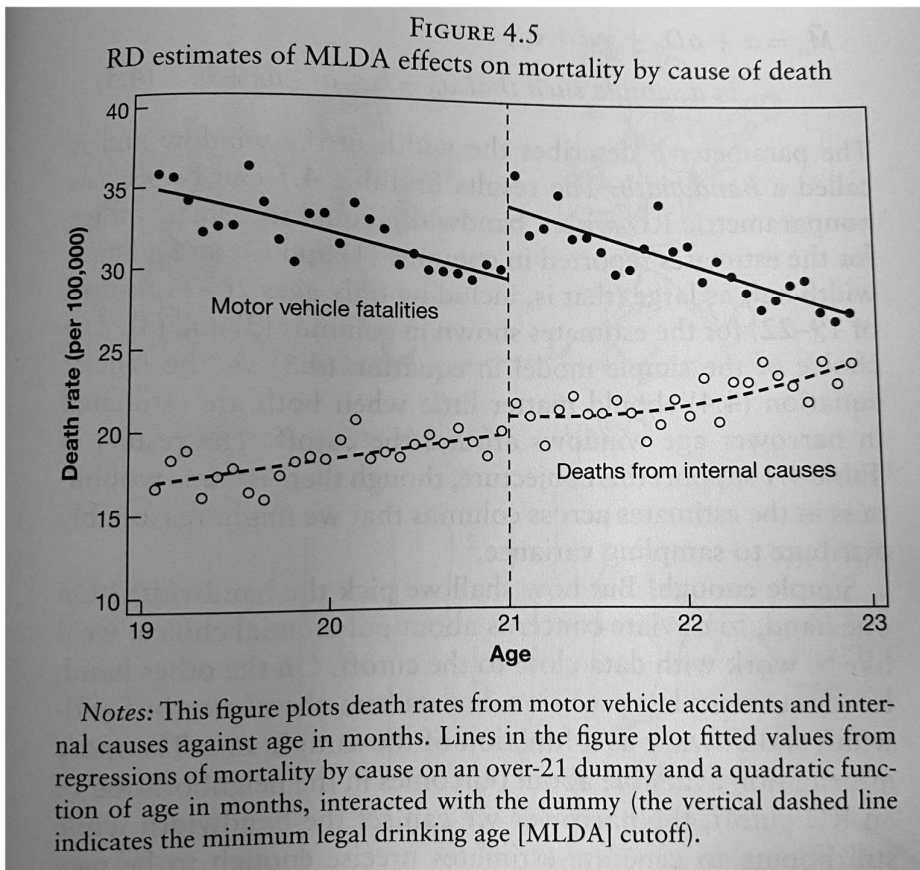
TABLE 4.1  
Sharp RD estimates of MLDA effects on mortality

Dependent variable	Ages 19-22		Ages 20-21	
	(1)	(2)	(3)	(4)
All deaths	7.66 (1.51)	9.55 (1.83)	9.75 (2.06)	9.61 (2.29)
Motor vehicle accidents	4.53 (.72)	4.66 (1.09)	4.76 (1.08)	5.89 (1.33)
Suicide	1.79 (.50)	1.81 (.78)	1.72 (.73)	1.30 (1.14)
Homicide	.10 (.45)	.20 (.50)	.16 (.59)	-.45 (.93)
Other external causes	.84 (.42)	1.80 (.56)	1.41 (.59)	1.63 (.75)
All internal causes	.39 (.54)	1.07 (.80)	1.69 (.74)	1.25 (1.01)
Alcohol-related causes	.44 (.21)	.80 (.32)	.74 (.33)	1.03 (.41)
Controls	age	age, age <sup>2</sup> , interacted with over-21	age	age, age <sup>2</sup> , interacted with over-21
Sample size	48	48	24	24

Notes: This table reports coefficients on an over-21 dummy from regressions of month-of-age-specific death rates by cause on an over-21 dummy and linear or interacted quadratic age controls. Standard errors are reported in parentheses.

- ・ Column(1)が年齢のみを制御した回帰式
- ・ Column(2)が年齢、年齢の二乗、処置変数とのInteractionを制御した回帰式。
- ・ 交通事故による死が多数を占め、因果効果は有意。
- ・ 自殺や外的要因による死亡も効果は小さいが有意。
- ・ 二つの回帰式による因果効果の違いは小さい。

- ・注目すべきは、内的要因（病気など）に対する因果効果は小さくかつ有意でない。
- ・直接的なアルコール中毒による効果も有意ではあるが効果自体は小さい。
- ・したがってこの表は、アルコールが原因で起こるような死亡に対する因果効果を示しているが、内的原因に関してあまり強い因果関係があるわけではない。



・表4.1のColumn(2)で推定した回帰式をプロットすると右図のようになる。  
（図はMVAと内的原因による死をプロットしたもの）

・紛らわしい非線形なトレンドのエビデンスはなく、明確なジャンプがMVAのほうに見られる。

・一方で内的原因のほうは明確なジャンプがみられない。

## 二つ目のモデル（ノンパラメトリックRD）

・二つ目のモデルは、境界線の付近のみに着目すればジャンプと非線形なトレンドを区別することは難しくないという事実を用いる。つまり、境界線付近のデータのみを用いて因果効果の推定を行う。

・欠点はデータ数が少なくなることで推定精度が落ちるところにある。

・推定方法は

$$\bar{M}_a = \alpha + \rho D_a + \gamma a + e_a$$

*in a sample such that  $a_0 - b \leq a \leq a_0 + b$*

$b$ は使用するデータの領域を決定し、バンド幅と呼ばれる。

29

TABLE 4.1  
Sharp RD estimates of MLDA effects on mortality

Dependent variable	Ages 19-22		Ages 20-21	
	(1)	(2)	(3)	(4)
All deaths	7.66 (1.51)	9.55 (1.83)	9.75 (2.06)	9.61 (2.29)
Motor vehicle accidents	4.53 (.72)	4.66 (1.09)	4.76 (1.08)	5.89 (1.33)
Suicide	1.79 (.50)	1.81 (.78)	1.72 (.73)	1.30 (1.14)
Homicide	.10 (.45)	.20 (.50)	.16 (.59)	-.45 (.93)
Other external causes	.84 (.42)	1.80 (.56)	1.41 (.59)	1.63 (.75)
All internal causes	.39 (.54)	1.07 (.80)	1.69 (.74)	1.25 (1.01)
Alcohol-related causes	.44 (.21)	.80 (.32)	.74 (.33)	1.03 (.41)
Controls	age	age, age <sup>2</sup> , interacted with over-21	age	age, age <sup>2</sup> , interacted with over-21
Sample size	48	48	24	24

Notes: This table reports coefficients on an over-21 dummy from regressions of month-of-age-specific death rates by cause on an over-21 dummy and linear or interacted quadratic age controls. Standard errors are reported in parentheses.

・したがってColumn(1),(2)の結果はバンド幅2年のノンパラメトリックRDといえる。

・Column(3),(4)はバンド幅を一年に縮めてそれぞれの回帰式を推定したもの。

・小さいバンド幅で推定されているときはモデル選択はあまり重要ではない。

30

# どのようにしてバンド幅を選ぶか？

- ・多項式による推定の技術的コストを減らしたい一方でデータ数が少ないことによる精度の低い推定は避けたい。
- ・したがって、データ数が十分に大きいときには小さいバンド幅で推定すればよい。
- ・実際は、バンド幅の決定も他者からの評価を要する。
- ・目的は、ただ一つの真のバンド幅を決めることでなく、むしろある特定のバンド幅によって得られた因果効果が偶然でないことを示すことである。

31

- ・今回の研究では、高次の説明変数を含めてもノンパラメトリックな推定法を使っても、因果効果の値はあまり変わらない。
- ・バンド幅を変えても因果効果の値はあまり変化しない。
- ・したがってRDによって解析されたMLDAの因果効果は本物の因果効果をとらえている。

32



# まとめ

- ・政策が適用される人とされない人の不連続性がRDDを適用することが可能であるような状況を作り出す。
- ・非線形なトレンドをもともと持っているのか、政策によってジャンプが生まれたのか見極める必要があり、高次の説明変数を含めてある程度制御できる。
- ・多項式で回帰する方法のほかに、バンド幅を定めて推定するノンパラメトリックRDと呼ばれる方法もある。