第3回　基礎ゼミ

回帰分析と因果効果分析

福田研究室　修士1年
城間　洋也

Outline

- イントロ
  因果効果分析の重要性
  因果効果の定義

- 回帰分析と因果効果分析
  回帰分析による因果効果推定＋検定
  推定された因果効果の信頼性（推定量の性質）
  回帰分析周辺のいろんな分析手法（一般化最小2乗法，操作変数法，パネルデータ分析）

- Rの便利な機能

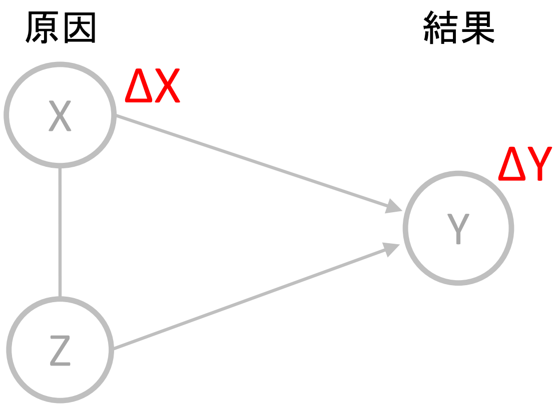
イントロ

因果効果分析の必要性

政策決定の場において、その政策が望ましい結果をもたらすという証拠（エビデンス）を科学的手法に基づいてより定量的に評価する重要性が高まっている

因果効果分析の目的 （Holland（1986））

1. 観察された結果に対する原因の究明
2. 観察された因果関係に対する因果メカニズムの究明
3. 観察された因果関係における因果効果の定量的評価



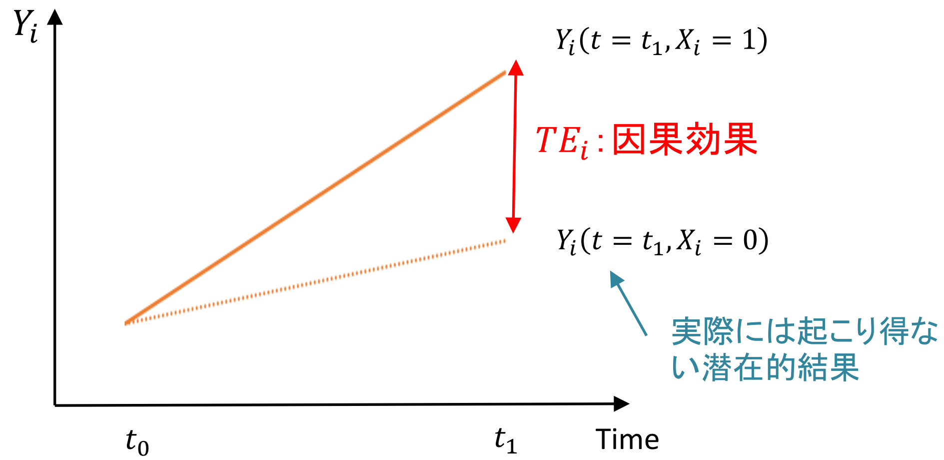
因果効果分析による政策評価

- 分析手法の妥当性・透明性
- 政策評価の精度

イントロ

<u>Rubinの因果モデル</u>（Rubin（1974））

　因果効果は潜在的結果を用いて図のように定義され，因果効果分析では各主体が受ける因果効果（TE_i）の期待値（平均因果効果ATE）を評価する



Y_i: 目的変数（結果），X_i: 政策変数（原因），t:時点

問題点：
- 潜在的結果は観測不可で因果効果は計算できない

潜在的結果を統計的特性を利用し補い因果効果を推計する

（ex）
- ランダム化比較試験（RCT）
- 回帰不連続デザイン（RD）
- マッチング法
- 差分の差分法（DD）

回帰分析と因果効果分析

回帰分析と因果関係

回帰分析　→ 目的変数*Y*と説明変数*X*の相関関係をモデル化+パラメータ推定

相関関係

因果関係
$X \rightarrow Y$

逆の因果, 同時性
$Y \rightarrow X, Y \leftrightarrow X$

共通要因
$C \rightarrow X,\ C \rightarrow Y$

偶然の相関

相関関係 ≠ 因果関係

回帰分析の推定式が*X*(原因)と*Y* (結果)の因果関係になるとは限らない

因果効果分析の注意点

- 目的変数と説明変数の時間的方向性　→ 実験的・準実験的手法, 時系列データ
- 他の要因の制御　→ 重回帰分析, 操作変数法...

回帰分析と因果効果分析

回帰モデルによる因果効果推定

回帰モデル:
$Y_i = \beta_0 + \beta X_i + \boldsymbol{\gamma C_i} + \boldsymbol{\varepsilon_i}$

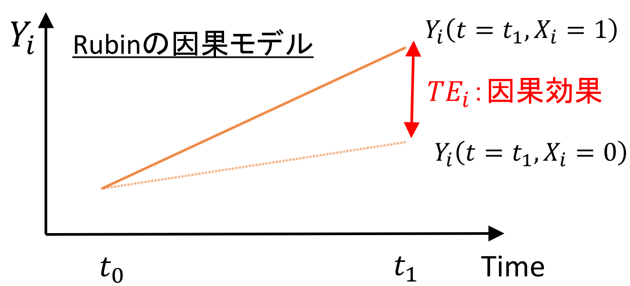
$\xrightarrow{\text{OLS推定}}$

推定結果:
$E[Y|X, \boldsymbol{C}] = \hat{\beta}_0 + \hat{\beta}X + \hat{\boldsymbol{\gamma}}\boldsymbol{C}$

X(原因)がΔX変化することがY(結果)に与える平均的な因果効果は

$$E[Y|X + \Delta X, \boldsymbol{C}] - E[Y|X, \boldsymbol{C}] = \hat{\beta}\Delta X$$

X以外の要因は一定としたときのYの期待値



有意性の検証

通常の回帰分析と同様にt検定により因果効果の有意性を確認できる
また，因果効果分析では回帰モデルの決定係数R^2が小さくても，目的の推定量($\hat{\beta}$)が一致性・不偏性・効率性を持つ推定量であればOK

一致性:サンプルサイズを大きくしたとき，推定量がある値に収束する性質
不偏性:推定量の期待値が真の値に等しい性質
効率性:推定量の分散が小さい性質

回帰分析と因果効果分析

推定量の信頼性

より信頼性の高い因果効果推計をするために一致性・不偏性・
効率性をもつ推定量を得ることが求められる

ガウスマルコフの定理

最小二乗推定量（OLS推定量）が一致性・効率性を持つための誤差項の条件

- 均一分散

（成立しないケース）
平均値データ分析
線形確率モデル

→

（対処法）
不均一分散頑健推定量を算出
一般化最小二乗法（GLS）

- 誤差項間に相関が無い

（成立しないケース）
時系列データ分析
パネルデータ分析

→

（対処法）
一般化最小二乗法（GLS）

- 説明変数と誤差項に相関関係がない

（成立しないケース）
内生性バイアス
欠落変数バイアス

→

（対処法）
説明変数を加える
操作変数法
パネルデータ分析

回帰分析と因果効果分析

一般化最小二乗法(GLS)

誤差項$\boldsymbol{u}$が不均一分散あるいは誤差項間に相関がある(自己相関)とき
→ OLS推定量より良い特性をもつ推定量がある

誤差項の共分散行列を$E(\boldsymbol{u}\boldsymbol{u}') = \sigma^2\boldsymbol{\Omega}$としたとき，$\boldsymbol{P}\boldsymbol{\Omega}\boldsymbol{P}' = \boldsymbol{I}$となる$\boldsymbol{P}$が存在する
この$\boldsymbol{P}$を用いて元の回帰モデルを変換しOLS推定することでより良い推定量が求まる

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u} \xrightarrow{\text{変換}} \boldsymbol{P}\boldsymbol{Y} = \boldsymbol{P}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{P}\boldsymbol{u} \xrightarrow[E[\boldsymbol{Pu}]=0,\ var(\boldsymbol{Pu})=\sigma^2\boldsymbol{I}]{\text{OLS推定}} \tilde{\beta} = (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{Y}$$

$\boldsymbol{\Omega}$にはモデルの誤差相関に応じて適当なものを使う

(ex)時系列相関

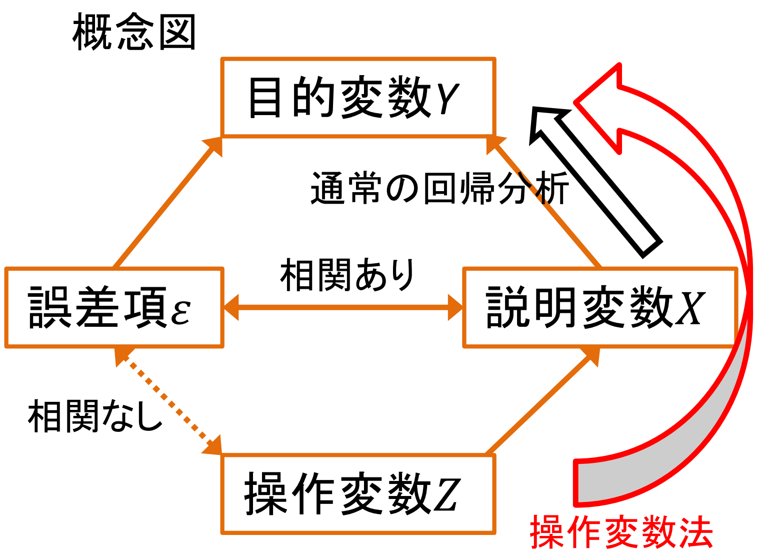
$$\boldsymbol{u}_{it} = \underbrace{\boldsymbol{\mu}_i}_{\text{主体特有の効果}} + \underbrace{\boldsymbol{\varepsilon}_{it}}_{\text{ランダムな誤差項}}$$

$$\boldsymbol{\Omega} = \begin{pmatrix} \sigma_\mu^2 + \sigma_\varepsilon^2 & \sigma_\mu^2 & \cdots & \sigma_\mu^2 \\ \sigma_\mu^2 & \sigma_\mu^2 + \sigma_\varepsilon^2 & \cdots & \sigma_\mu^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\mu^2 & \sigma_\mu^2 & \cdots & \sigma_\mu^2 + \sigma_\varepsilon^2 \end{pmatrix}$$

回帰分析と因果効果分析

操作変数法

→　説明変数と誤差項に相関がある場合の回帰分析



（ex）

- 学歴が賃金に与える影響の分析
  操作変数：両親の学歴
- 出席率が成績に与える影響の分析
  操作変数：通学時間
- 母親の喫煙が子供の出生時の体重に与える影響
  操作変数：タバコの税率

2段階最小二乗法

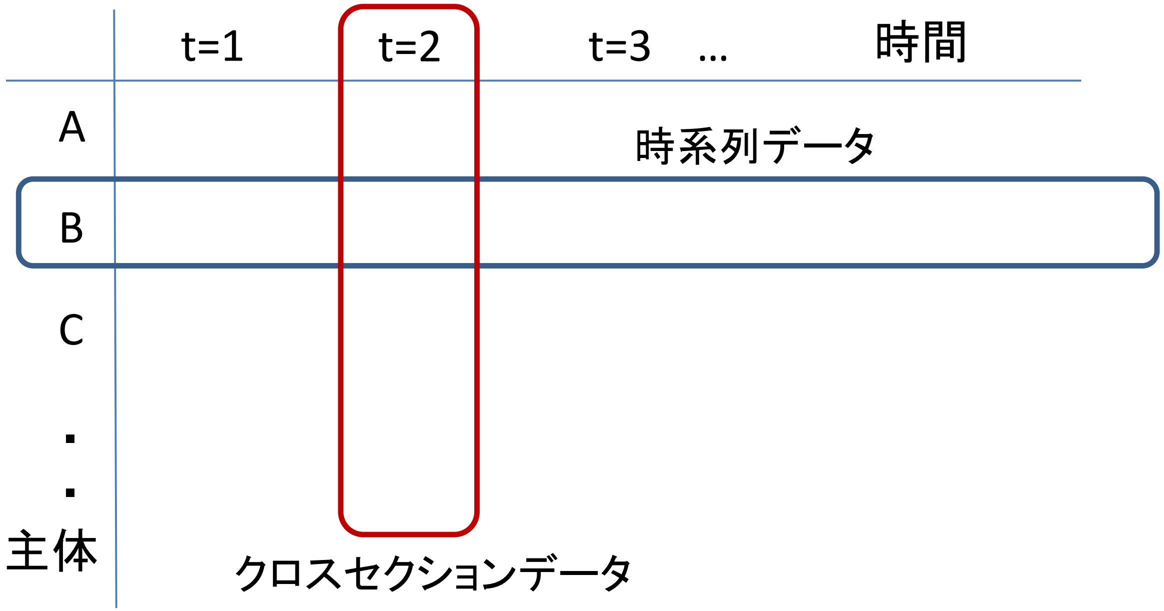
$Z \to X$　の回帰　→　$\hat{X} \to Y$　の回帰

一致推定量が得られる

回帰分析と因果効果分析

パネルデータ分析

パネルデータとは…



パネルデータ ＝ 時系列データ ＋ クロスセクションデータ

↓

複数主体の時系列変化を観察できるため因果効果分析に適している

回帰分析と因果効果分析

パネルデータ分析

（メリット）

- データ量豊富
- 主体特有の時間不変の効果を制御可能

通常の回帰モデル

$$Y_i = \beta_0 + \beta X_{it} + \boldsymbol{\gamma C_i} + \boldsymbol{\gamma C'_i} + \varepsilon_{it}$$

観測不能

$\boldsymbol{C'_i}$が欠落変数となり推定量にバイアスが生じる

パネル回帰

$$Y_{it} = \beta_0 + \beta X_{it} + \mu_i + \varepsilon_{it}$$

$\boldsymbol{\gamma C_i} + \boldsymbol{\gamma C'_i}$を主体固有のダミー変数として推定することで観測不可能な効果まで制御できる(固定効果推定)

結果としてより信頼性の高い因果効果$\hat{\beta}$の推定が可能

Rの便利な機能

回帰分析の関数

lm():通常のOLS推定
glm():GLS推定

便利なパッケージ

robust:不均一分散頑健推定量を算出してくれるパッケージ

plm:パネルデータ分析用のパッケージ
→ 通常のOLS，固定効果推定，変量効果推定（一般化最小二乗法）ができる

stargazer:回帰分析の結果をかっこいい表にまとめてくれるパッケージ
Texコードで出力されます