Mistape Chapter8: Panel Data

牧野泰季
山岸翔一朗
伊藤璃杏

Overview

1- Panel Data

- パネルデータとは何か？
- パネルデータでできること

2- Panel Data Analysis

- Pooled OLS model プールしたOLS推計モデル：欠落変数バイアスの問題
- Fixed Effects model　固定効果モデル
- Random Effects model　変量効果モデル

3- Panel Data Analysis- in practice-

- Panel Data Analysis_ Example

4- Appendix

Panel Data

パネルデータとは何か？

データ構造の比較

	一時点のデータ At one point in time	複数時点のデータ At many points in time
変数は1つ Data from one unit	-	時系列データ Time series data
変数は複数 Data from many units	クロスセクションデータ Cross sectional data	パネルデータ Panel data (/ Time series cross section)

パネルデータとは何か？

データ構造の比較

- **横断面データ/クロスセクションデータ**

(Cross-section data)

ある時点における場所・グループ別などに記録した、複数の項目（個体）を集めたデータ

e.g.,) ID, 性別, 年齢, 雇用体系, 配偶者の有無 etc.

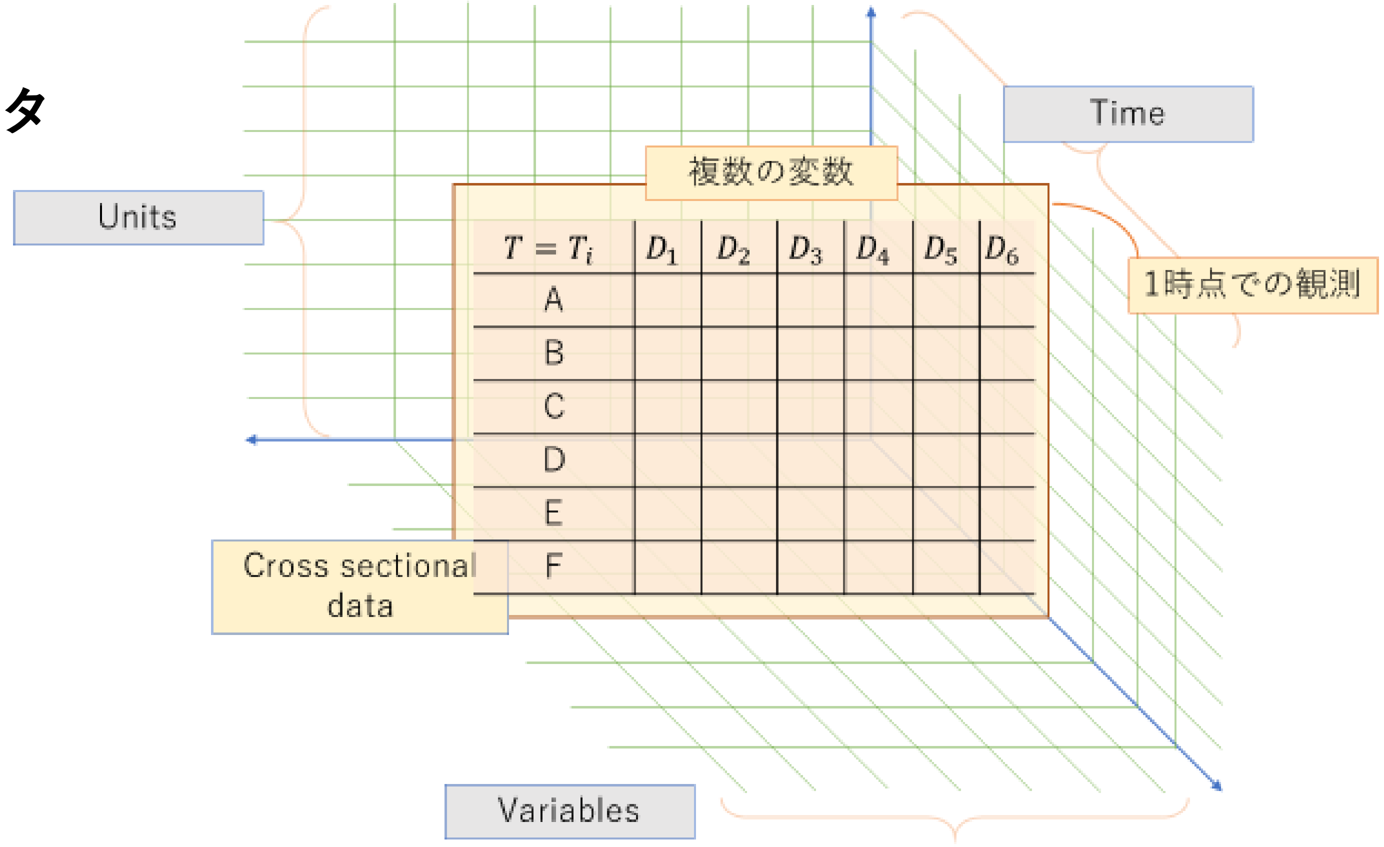
・横断調査：観察時点が一点（固定）で、複数の項目（個体）を観察する

$T = T_i$	D_1	D_2	D_3	D_4	D_5	D_6
A						
B						
C						
D						
E						
F						

パネルデータとは何か？

データ構造の比較

- **横断面データ/クロスセクションデータ (Cross-section data)**

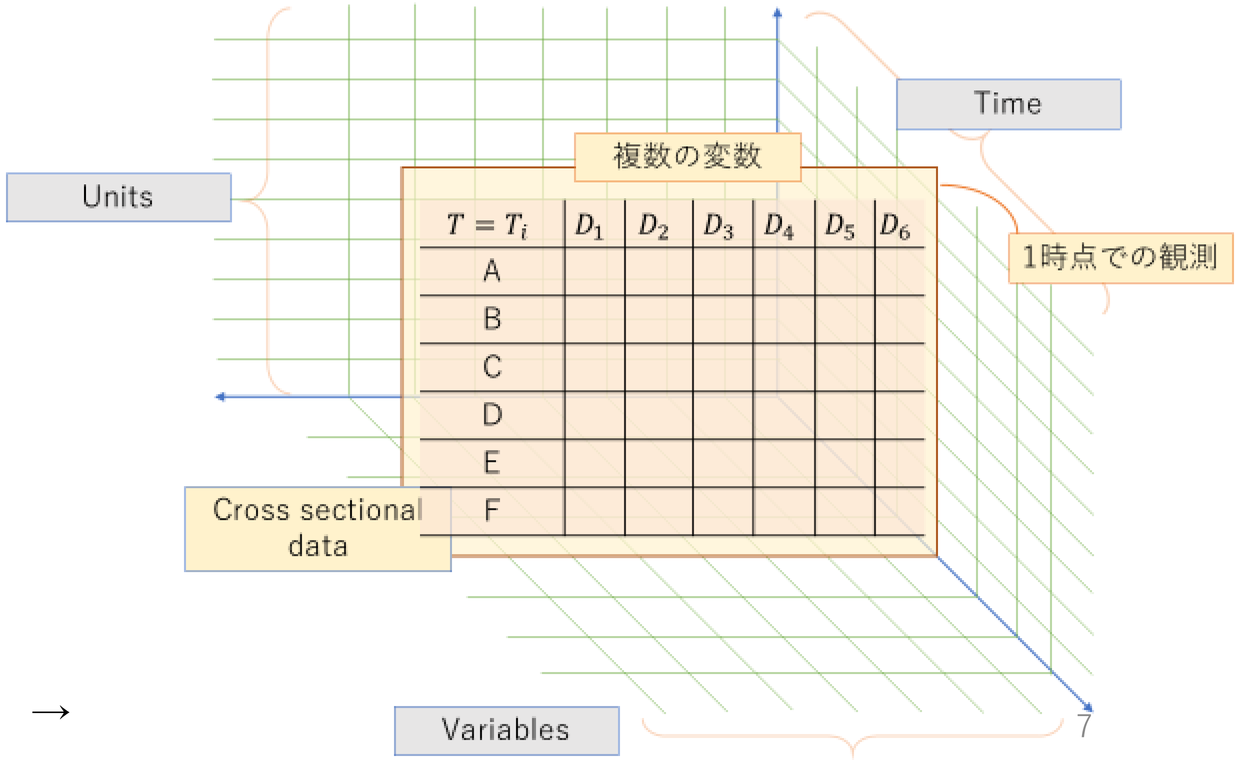


パネルデータとは何か？

データ構造の比較

- **横断面データ/クロスセクションデータ(Cross-section data)**

クロスセクションデータのイメージ　→



パネルデータとは何か？

データ構造の比較

- **時系列データ（Time-series data）**

  時間的に変化した情報を持つデータ

  e.g.,) トレンド, 季節変動, 循環変動, 不規則変動 etc.

  ・時系列調査：一つの項目について継続的に観察する

$D = D_i$	T_1	T_2	T_3	T_4	T_5	T_6
A						
B						
C						
D						
E						
F						

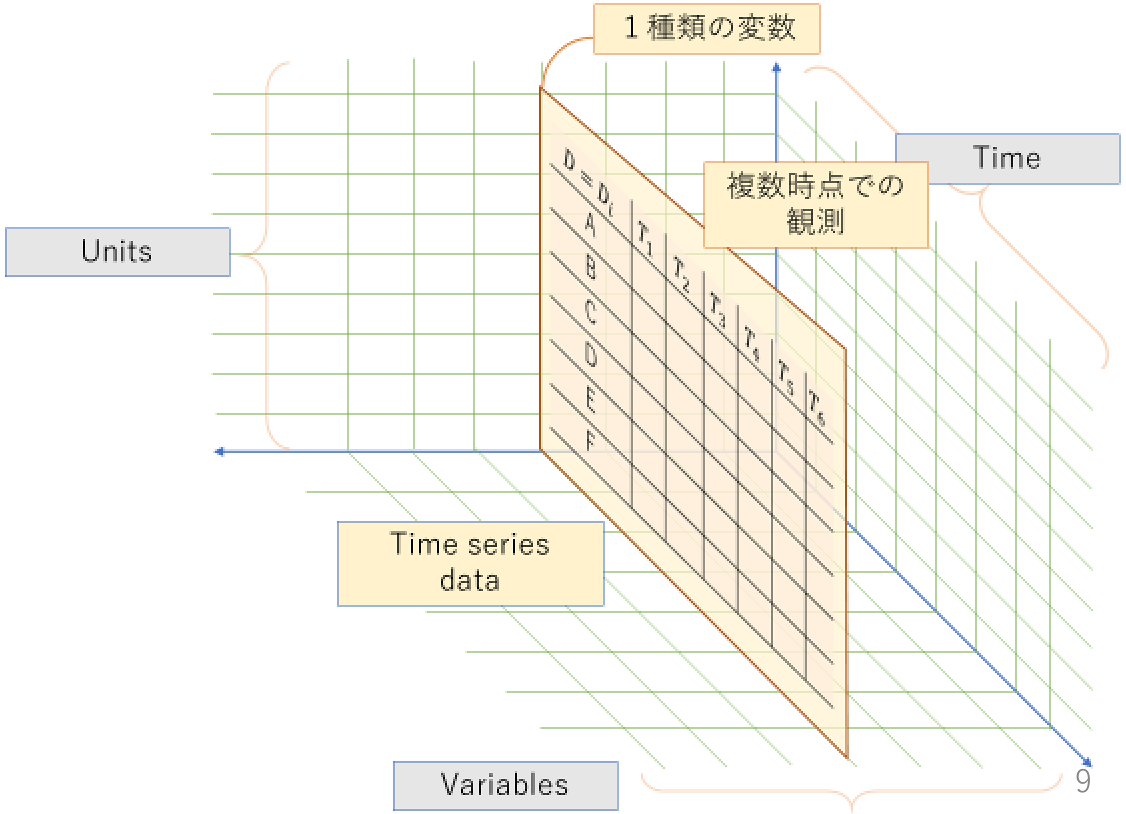
↑時系列データのイメージ

パネルデータとは何か？

データ構造の比較

- **時系列データ(Time-series data)**

時系列データのイメージ　→



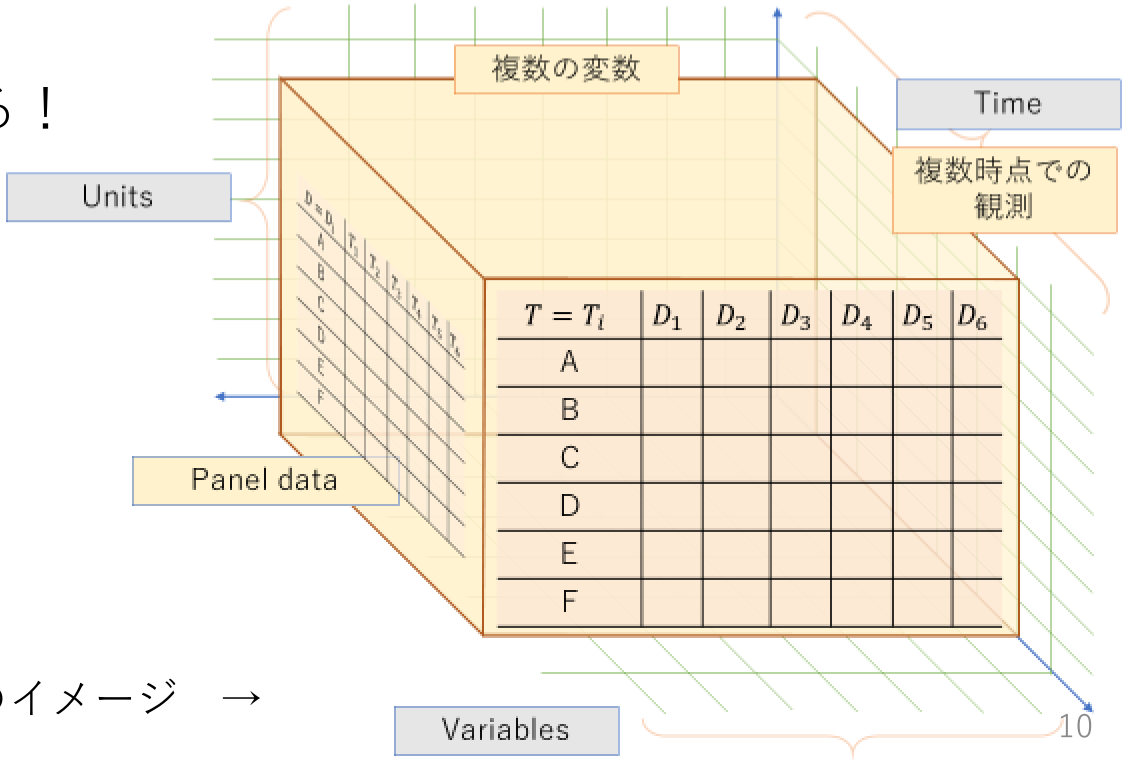
パネルデータとは何か？

データ構造の比較

- **時系列データ(Time-series data)**

各個体を各項目（変数）ごとに

継続的に追跡したデータが得られる！



時系列データのイメージ　→

パネルデータとは何か？

データ構造の比較

- **パネルデータ（Panel data）**

  同一のサンプルに対して複数の項目を複数時点にわたって調査したデータ

  →　クロスセクションデータ・時系列データの特徴を併せ持つデータ、と言える！

  ・パネル調査：「同一の時点で複数個体に対して行う調査」を複数時点にわたって繰り返し行う

複数項目

		D_1	D_2	D_3
A	T_1			
B	T_1			
C	T_1			
A	T_2			
B	T_2			
C	T_2			
A	T_3			
B	T_3			
C	T_3			

パネルデータのイメージ↑

パネルデータでできること

- ✓ これまで観察不可能であった潜在変数を推定したり, 経済主体のダイナミックな変動などを理解することもできるようになる
- ✓ **観察されない個々の異質性**によって引き起こされる**欠落変数バイアス**の一部を取り除くことができる

パネルデータでできること

●**観察されない個々の異質性（Unobserved heterogeneity）：**

直接観察できない個人特性のこと。

性格、能力etc.

- これによって引き起こされるのが**欠落変数バイアス (Omitted variable bias)**

●**欠落変数バイアス (Omitted variable bias)：**

介入や処置と相関の考えられる、本来コントロールすべき(回帰式に加えるべき)

共変量をモデルから取り除いてしまうことで生じるバイアス

パネルデータでできること

パネルデータ分析に向いている例：
「夫婦の労働時間の変化が家事分担比率に及ぼす影響」

仮説：

妻の労働時間：増加
夫の労働時間：一定　とすると、

夫が家事を負担する割合は増加する

パネルデータでできること

パネルデータ分析に向いている例：
「夫婦の労働時間の変化が家事分担に及ぼす影響」

横断調査だと、、

妻の平均労働時間1hr/日

	$w_w = 1$	$w_w = 2$	$w_w = 3$	$w_w = 4$
2020	0.88	0.87	0.78	0.61

妻の家事負担割合

妻の労働時間が短い夫婦と長い夫婦という異なった個体において家事分担比率を比べることになる

パネルデータでできること

パネルデータ分析に向いている例：
「夫婦の労働時間の変化が家事分担に及ぼす影響」

Q. これは本当に妻の労働時間の増減による効果なのか？

- ✓ ジェンダー平等志向が強い夫婦では妻が長く働き, かつ夫が家事を多く負担している可能性がある

<u>夫婦の個体特性（ジェンダー平等志向）に起因する効果</u>も含まれているのでは？

パネルデータでできること

パネルデータ分析に向いている例：
「夫婦の労働時間の変化が家事分担に及ぼす影響」

パネルデータ分析だと、、

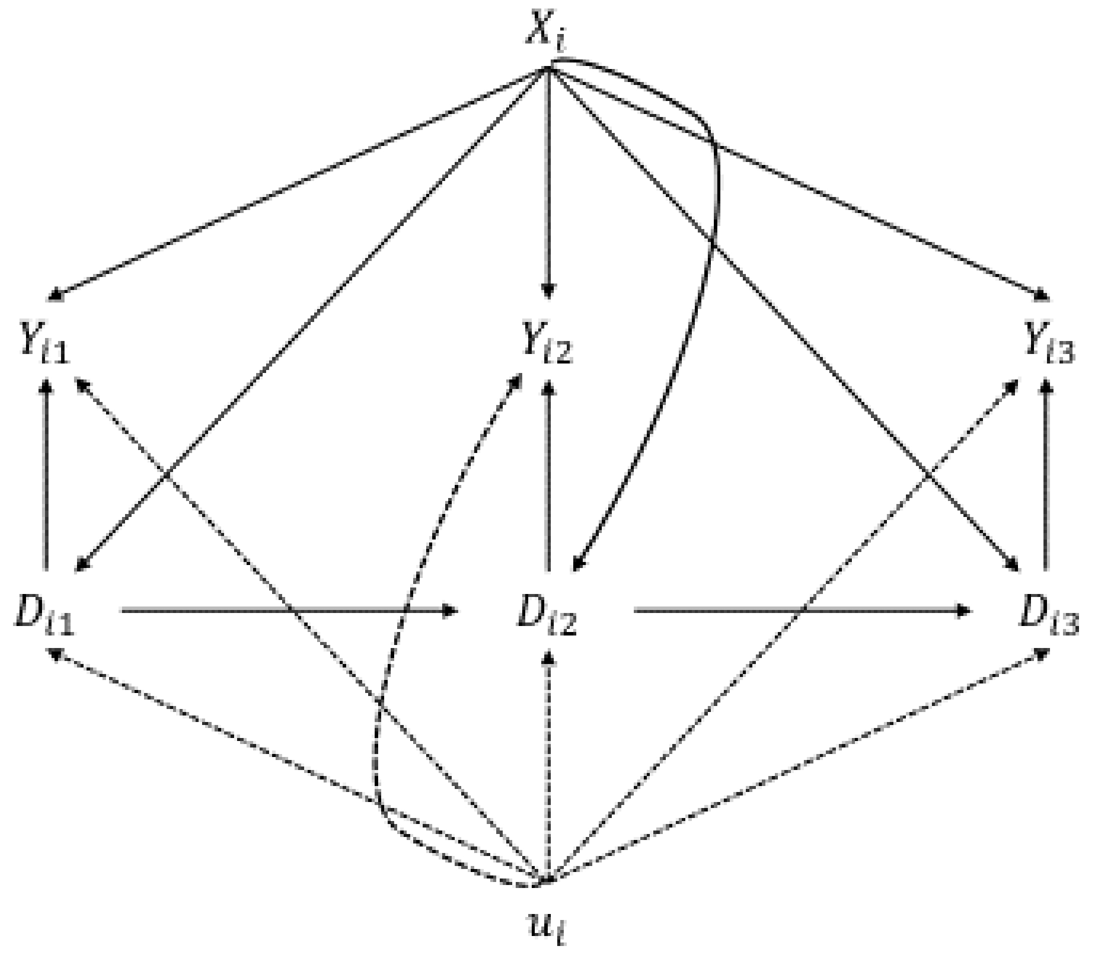
	A	B	C	D
2015	0.91	0.61	0.48	0.85
2020	0.88	0.61	0.47	0.76

個々の夫婦個体内の効果の平均値(平均的因果効果)を推定するため、観察されない個々の異質性に起因するバイアスは取り除かれている。

Panel Data Analysis

DAGによる整理

Y_{it} ：t期における、標本iの被説明変数
D_{it}：t期における、標本iの説明変数
X_i ：標本iに固有の効果（観測可能）
u_i ：標本iに固有の効果（観測不可能）

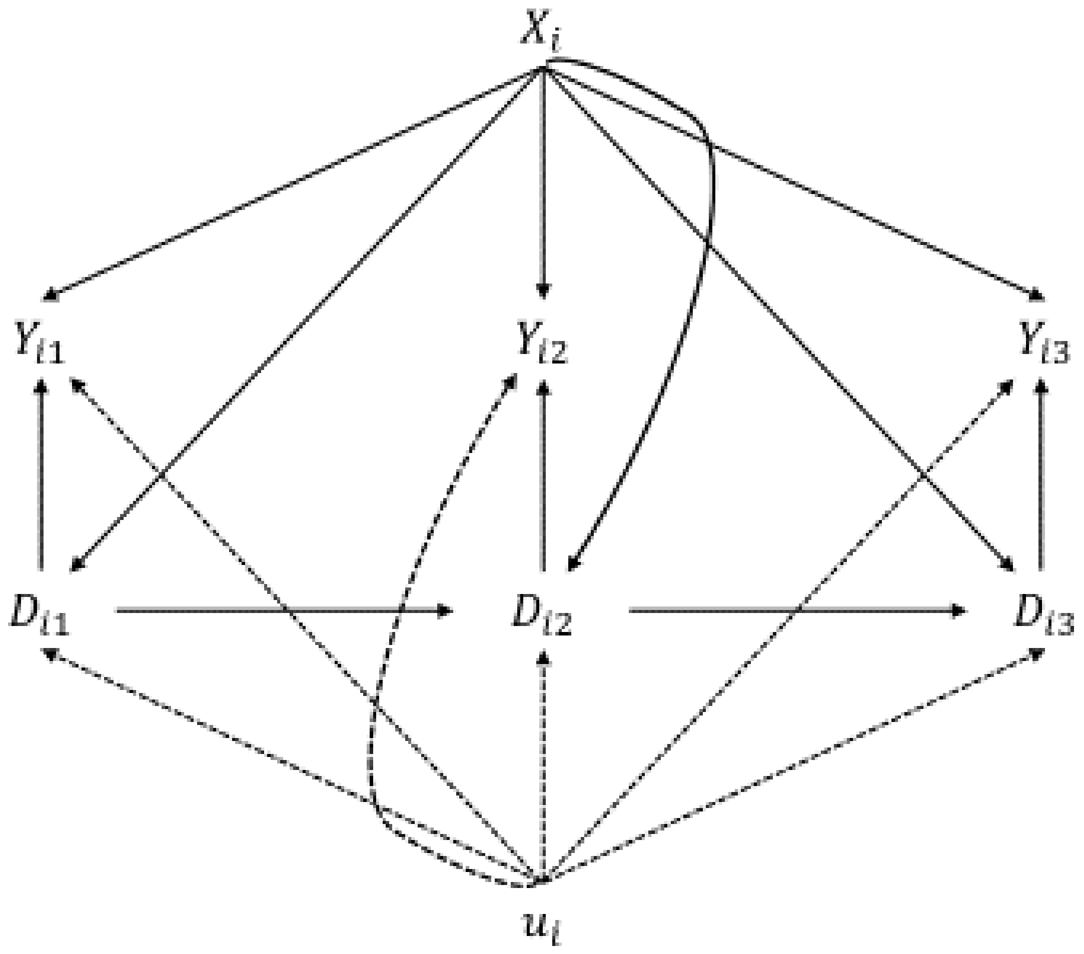


DAGによる整理

Y_{it} ：t期における、標本iの被説明変数
D_{it} ：t期における、標本iの説明変数
X_i ：標本iに固有の効果（観測可能）
u_i ：標本iに固有の効果（観測不可能）

① D_{i1}はY_{i1}と次の期間の推定値D_{i2}の両方に影響を与える

$$Y_{i1} \leftarrow \mathrm{D}_{i1} \rightarrow \mathrm{D}_{i2}$$

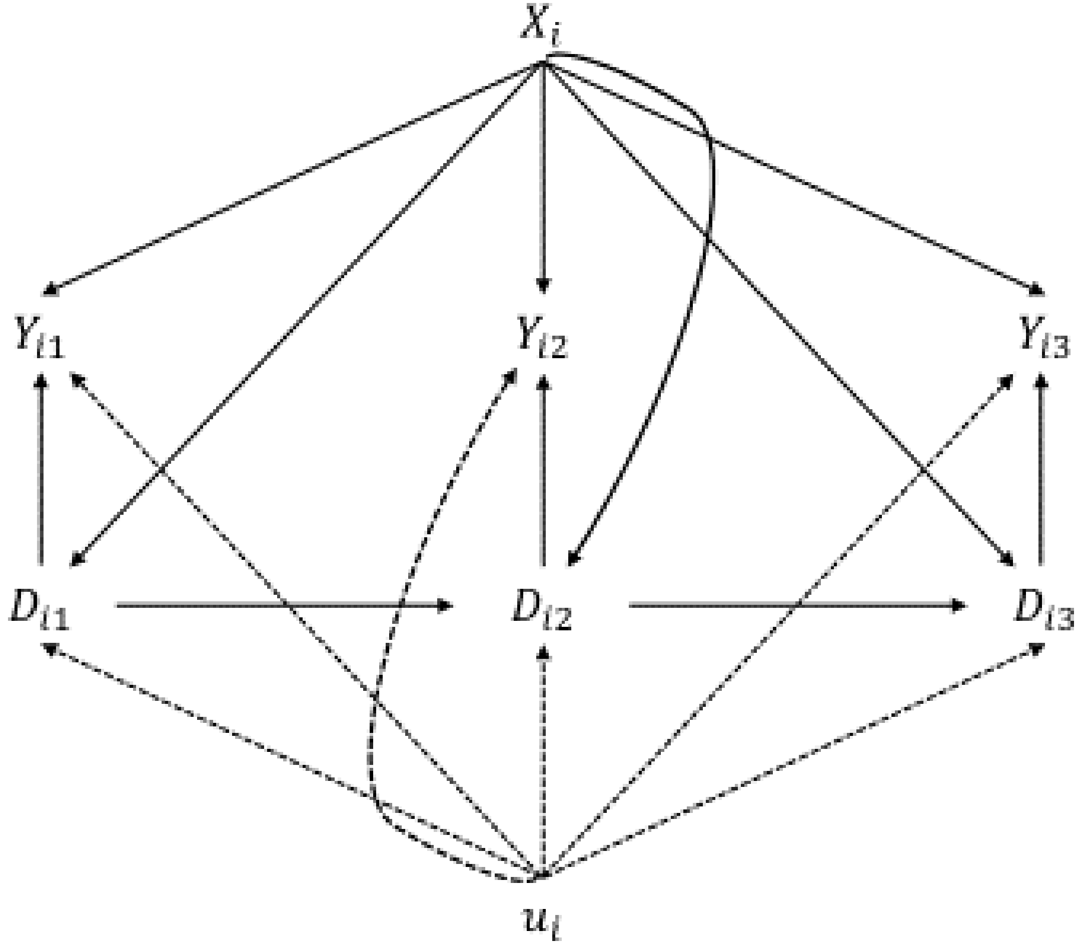


DAGによる整理

Y_{it} ：t期における、標本iの被説明変数
D_{it} ：t期における、標本iの説明変数
X_i ：標本iに固有の効果（観測可能）
u_i ：標本iに固有の効果（観測不可能）

②観察されない交絡変数u_iは
すべてのYとDの値を決める

- u_i "は観測されずに回帰分析の誤差項 (structural error term)に含まれる
＝ D_{it} "は内生的

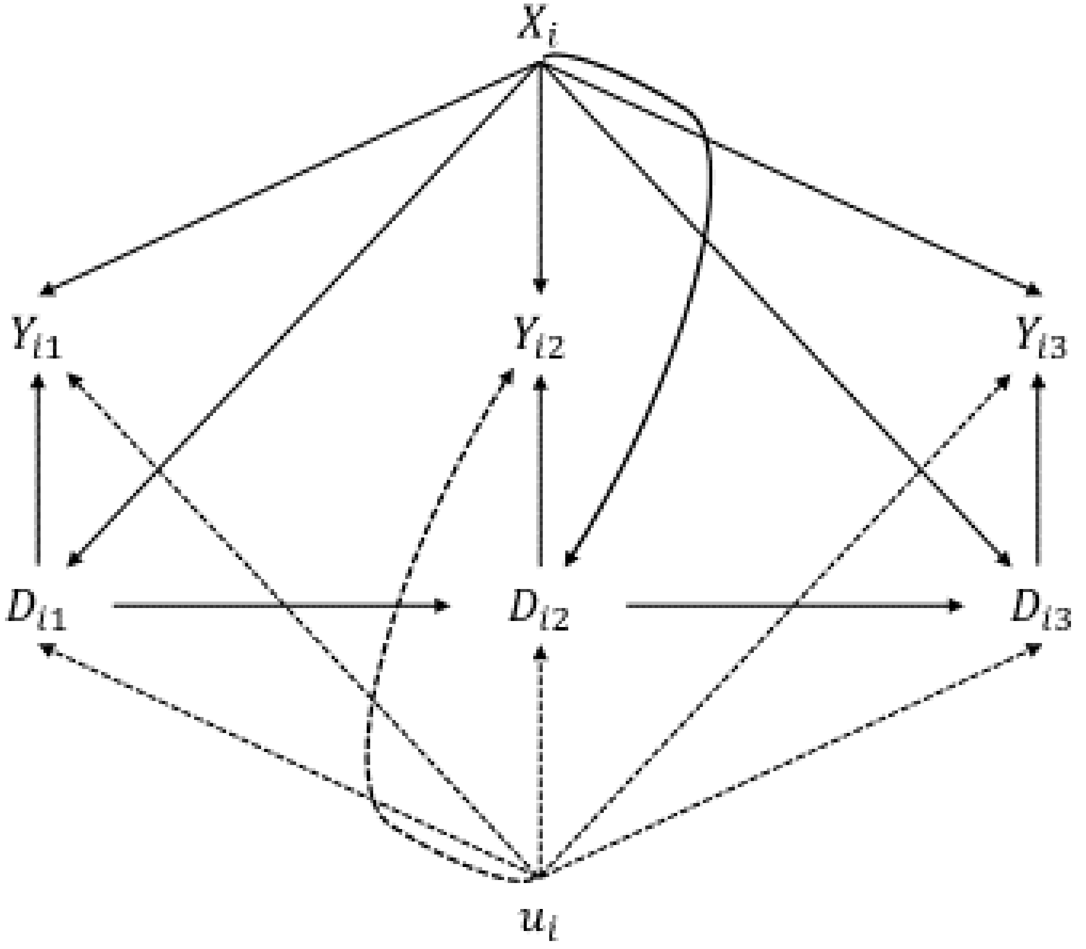


DAGによる整理

Y_{it} ：t期における、標本iの被説明変数
D_{it} ：t期における、標本iの説明変数
X_i ：標本iに固有の効果（観測可能）
u_i ：標本iに固有の効果（観測不可能）

③時間によって変化する、観察不可能な交絡因子（time-varying unobserved confounder）はない

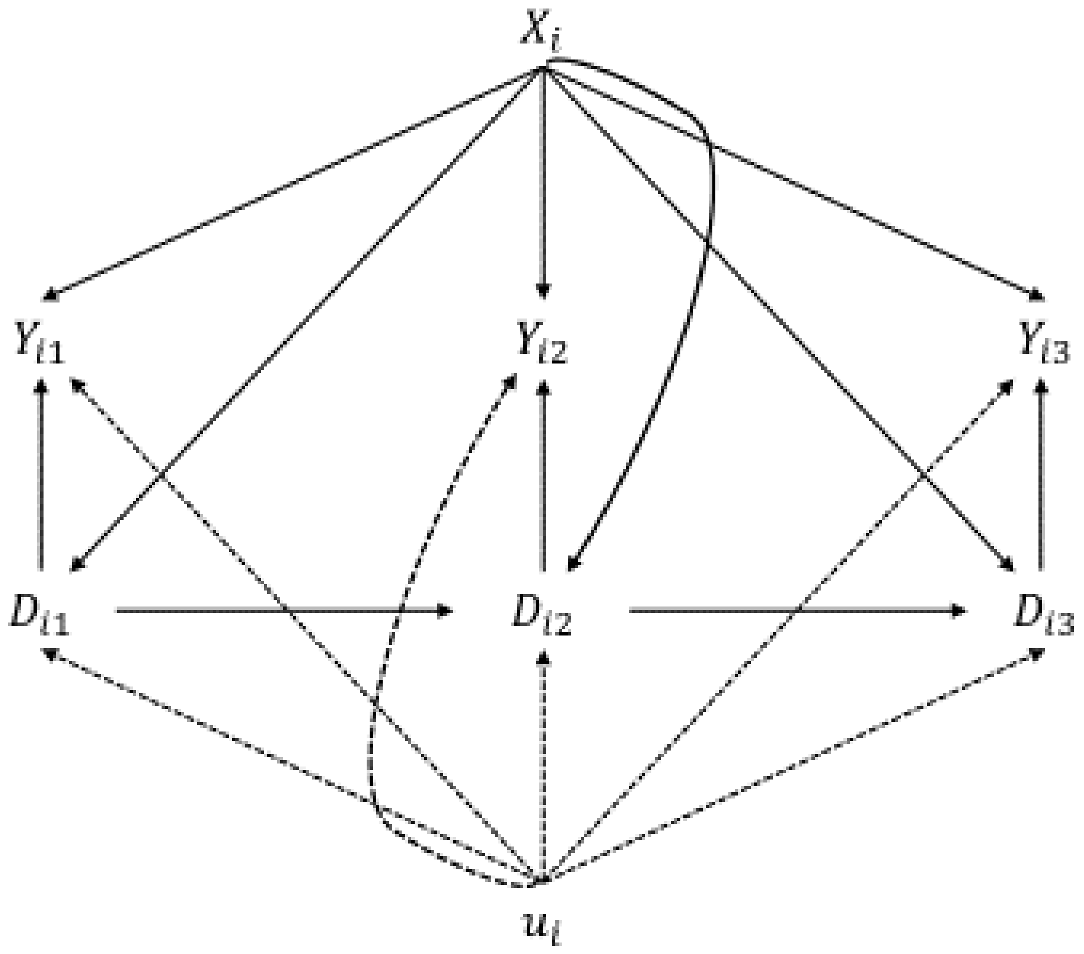
- 唯一の交絡因子はu_iである、これを観察されない異質性（unobserved heterogeneity） と呼ぶ



DAGによる整理

Y_{it} ：t期における、標本iの被説明変数
D_{it} ：t期における、標本iの説明変数
X_i ：標本iに固有の効果（観測可能）
u_i ：標本iに固有の効果（観測不可能）

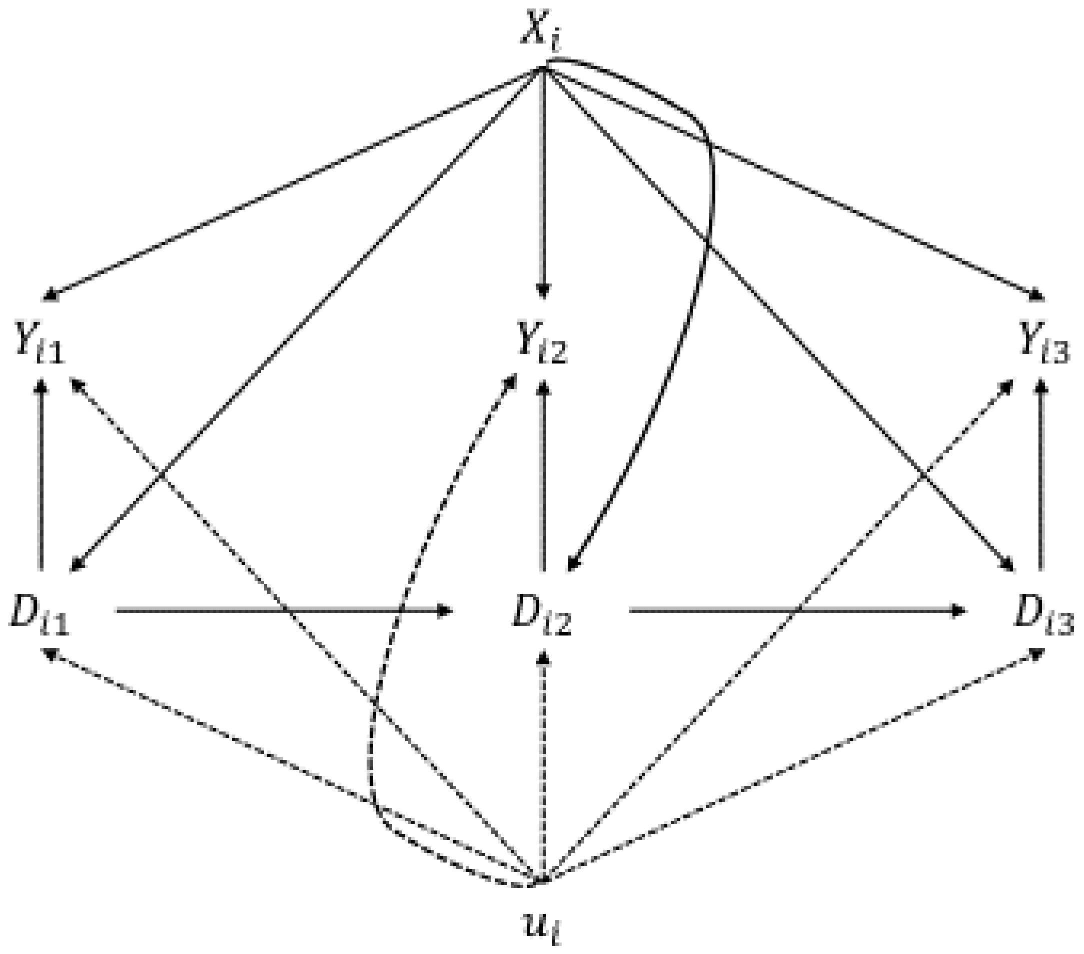
④過去の結果は現在の結果に直接的な影響を及ぼさない（ Y_{t-1}はY_tに直接的な影響はない）



DAGによる整理

Y_{it} ：t期における、標本iの被説明変数
D_{it} ：t期における、標本iの説明変数
X_i ：標本iに固有の効果（観測可能）
u_i ：標本iに固有の効果（観測不可能）

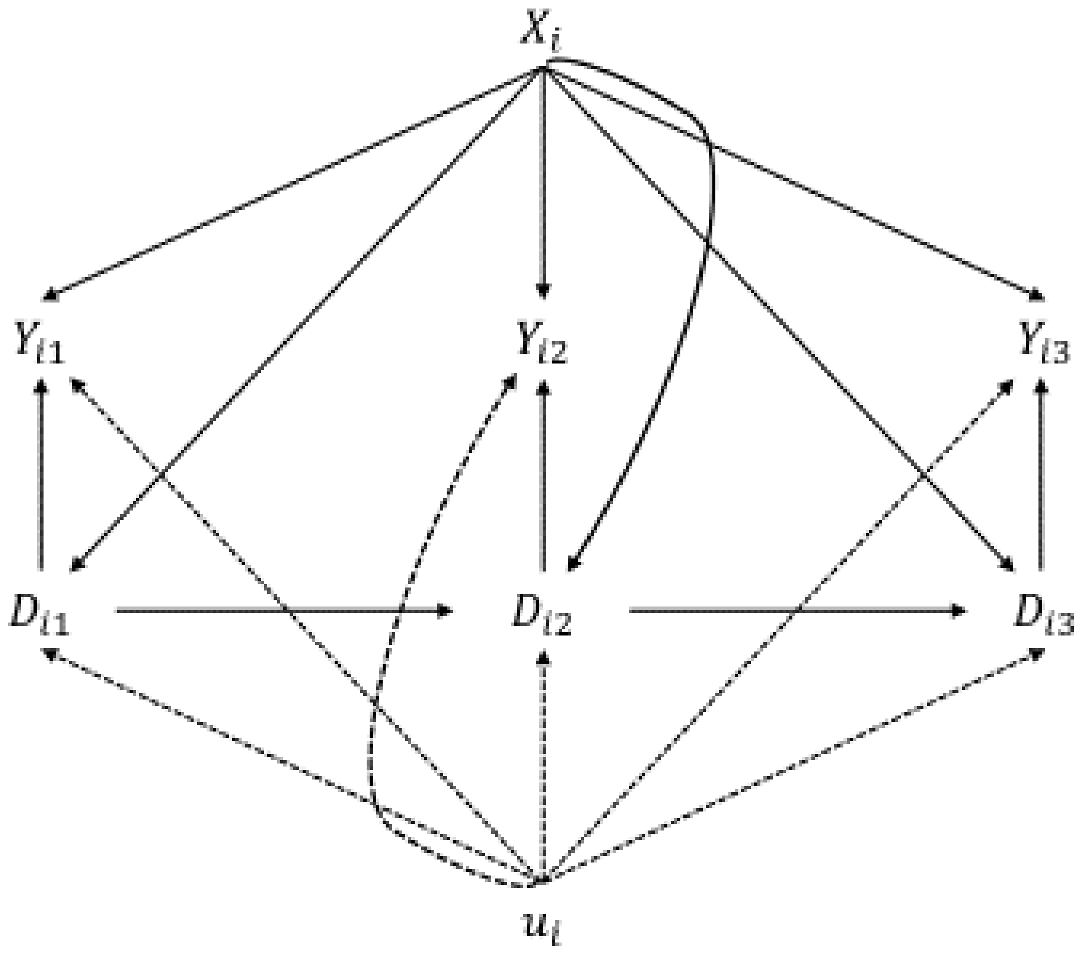
⑤過去の結果は現在の処置に直接的な影響を及ぼさない（ Y_{t-1}はD_tに直接的な影響はない）



DAGによる整理

Y_{it}　：t期における、標本iの被説明変数
D_{it}：t期における、標本iの説明変数
X_i　：標本iに固有の効果（観測可能）
u_i　：標本iに固有の効果（観測不可能）

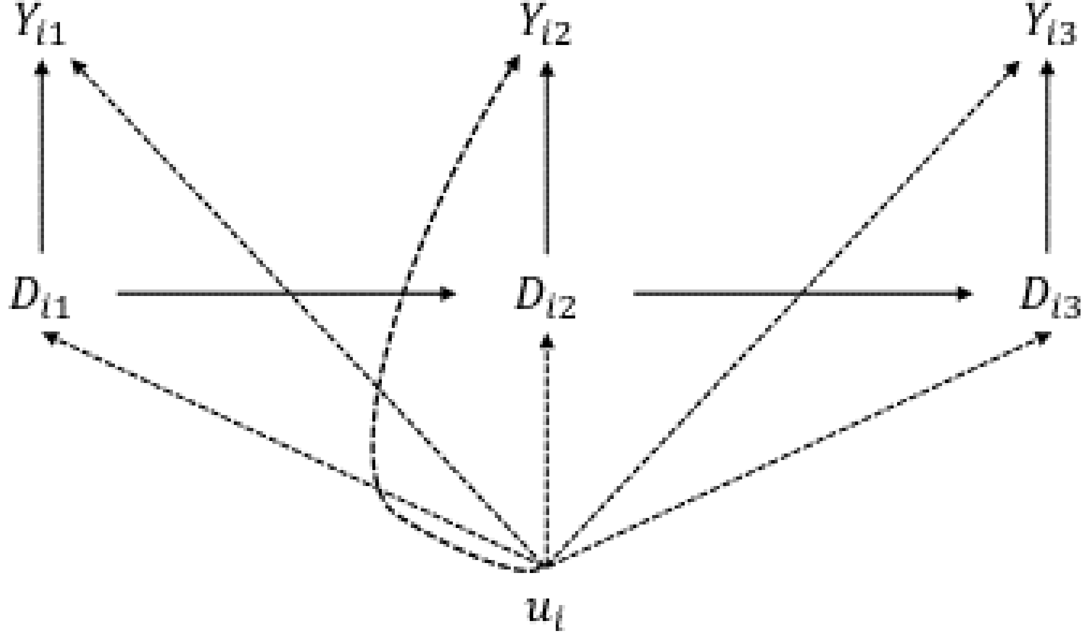
⑥過去の処置は現在の結果に直接的な影響を及ぼさない（ D_{t-1}はY_tに直接的な影響はない）



DAGによる整理

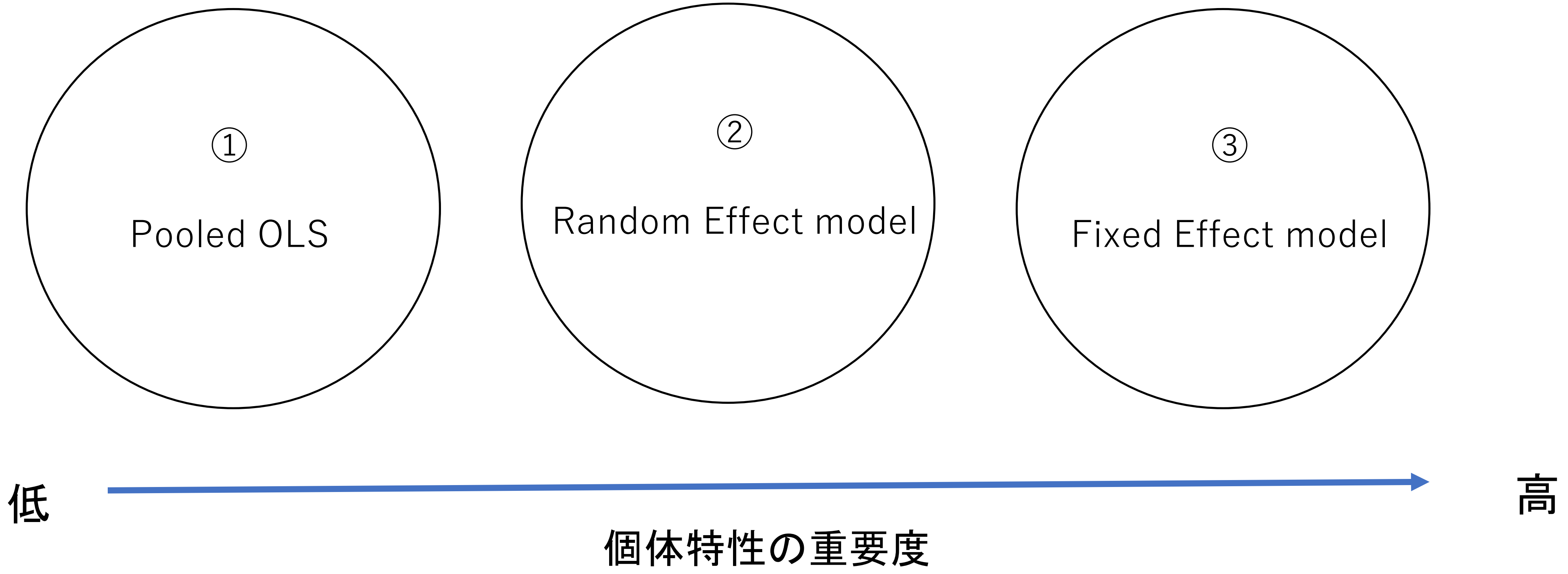
※Mixtapeでは、理論部分の説明においてX_iを除いて回帰分析を行っている為、ここでもそれに倣います。（X_iを含んでいてもやることと結果はほぼ同じですが、納得いかなければAppendixへ。）

※これ以降の計算はすべて行列計算ですが、
普通の一次式と見ていただいても特に問題ありません。数学苦手な方もご安心下さい。
（行列の場合の各項の姿を見たい方もAppendixへ。）



パネルデータを分析しよう

- パネルデータの分析方法は主に3つある

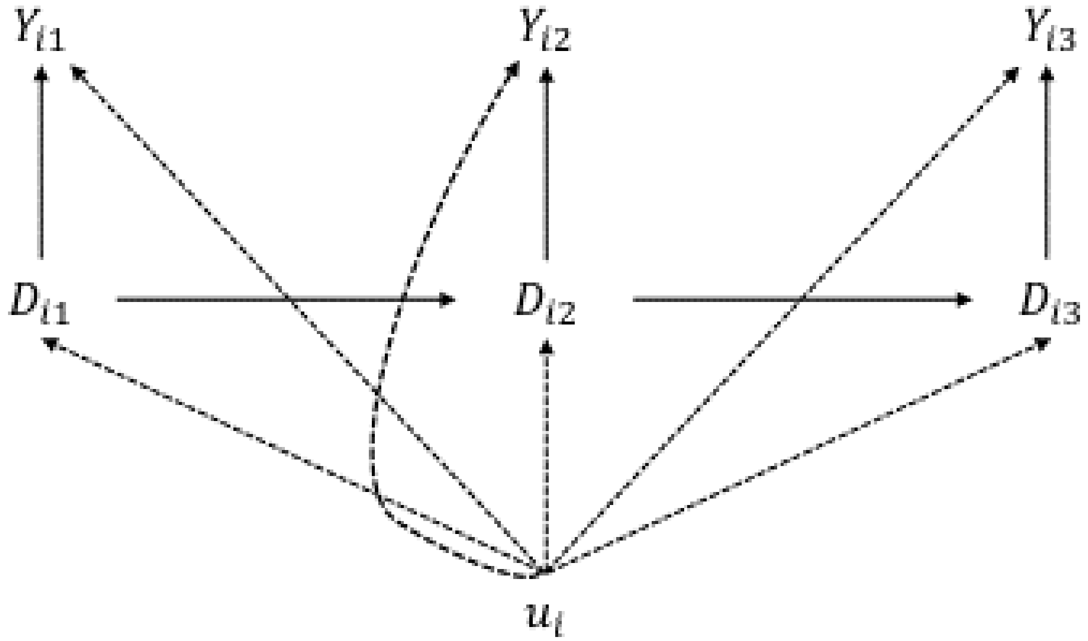


パネルデータを分析しよう

例：学校教育が収入に与える影響

$$Y_{it} = D_{it}\delta_i + u_i + \varepsilon_{it}$$

Y_{it} ：t期における、iの収入(log)
D_{it}：t期における、iの受ける学校教育
u_i ：iの知性(など、遺伝的要因)



イメージ：

$$\log(earning_{it}) = schooling_{it}\delta_i + intelligence_i + \varepsilon_{it}$$

用語：u_i …unobserved heterogeneity
ε_{it}…idiosyncratic error

個体固有効果 … individual specific effects

パネルデータを分析しよう

(Note)
Y_{it} ：$\log(earning_{it})$
D_{it} ：$schooling_{it}$
u_i ：$intelligence_{it}$

まずは素直にOLSで。

$$Y_{it} = D_{it}\delta_i + u_i + \varepsilon_{it}$$

u_iは非観測誤差（時間を通じて個人に一定の効果）

●集計最小二乗法(Pooled OLS)

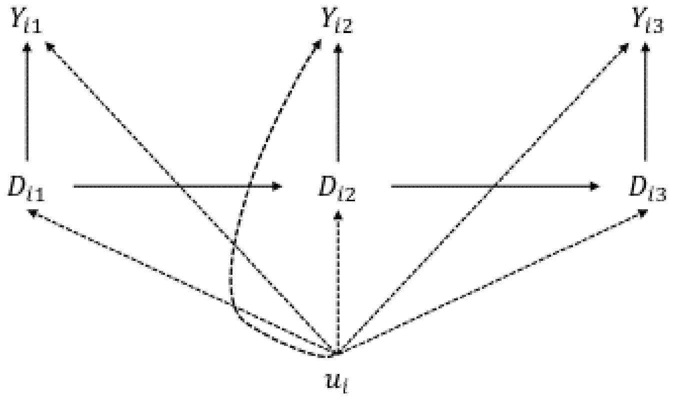
・データをCross-sectional なものと仮定して分析

要は、D_{it}とu_iが独立であることを仮定して分析

（u_iがジャマなので、もともと無かったことにした、

$$Y_{it} = D_{it}\delta_i + \eta_{it}$$

=> 係数の推定はできる！…けど、合ってる？



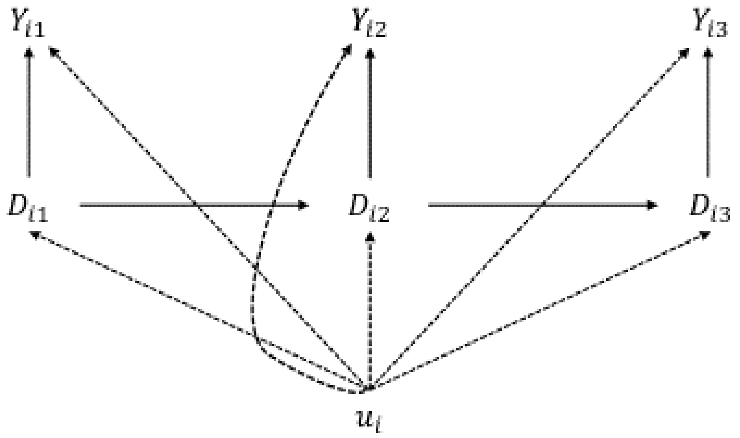
パネルデータを分析しよう

欠落変数バイアスの問題

就学することD_{it}は時間経過で変化しない個体特性u_iの影響を受けていると考えられているため、これを統制しないと欠落変数バイアス (omitted variable bias) が発生してしまう。

しかし、
D_{it}とη_{it}に相関がないことを仮定してしまうと、全ての時間軸において、u_iと就学することD_{it}に相関がないと仮定することになる。

→個体特性と就学することに相関がない、とするのは不自然



パネルデータを分析しよう

(Note)
Y_{it} : $\log(earning_{it})$
D_{it} : $schooling_{it}$
u_i : $intelligence_{it}$

欠落変数バイアスの問題

$$Y_{it} = D_{it}\delta_i + \eta_{it} \quad \dots \quad (\eta_{it} = u_i + \varepsilon_{it})$$

・実際には、D_{it} と u_i は独立ではなかった

$$\widehat{\delta_i} = \delta_i + \frac{Covariance(D_{it}, u_i)}{Variance(D_{it})}, \qquad Covariance(D_{it}, u_i) \neq 0$$

↑　ココの分、欠落変数バイアスが生じる　=>　ズレる…

パネルデータを分析しよう

不均一分散に関して頑健な誤差の問題

●不均一分散に関して頑健な誤差（Heteroscedastic Robust Standard Errors）
残差の均一性の仮定に対して、頑健な標準誤差の推定を行う

頑健性のある標準誤差（Robust Standard Error）
＊頑健性のある・ロバストな・（robust）：必要としている条件または仮定を少々満たしていないようなデータにおいてもほぼ妥当な結果を与える

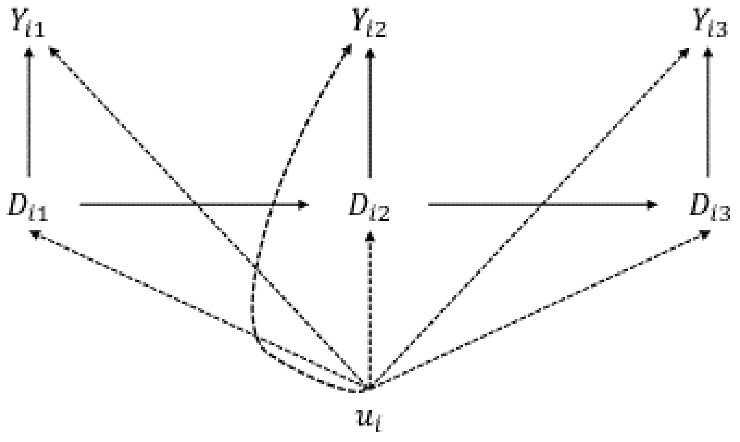
パネルデータを分析しよう

不均一分散に関して頑健な誤差の問題

u_iは各t期それぞれに存在するため、 $\eta_{it} \equiv (c_i + \epsilon_{it})$と iとが系列相関を持っているといえる。

●系列相関（serial correlation):
回帰分析において、誤差項（残差、実測値と理論値の差）に自己相関があること

●自己相関(autocorrelation)：
ある変数の系列中において、個々の変数値が一定間隔を置いて相関を持つこと。元データと時間をずらしたデータとの相関。



パネルデータを分析しよう

不均一分散に関して頑健な誤差の問題

$$Var(\widehat{\delta}_i) = \frac{\sum_{t=1}^{T}(D_{it} - \overline{D_i})^2 Var(\eta_{it}|D_{it})}{(\sum_{t=1}^{T}(D_{it} - \overline{D_i})^2)^2}$$

u_iの存在による、η_{it}の自己相関により、不均一分散に関して頑健な誤差が過少に推定されやすい

パネルデータを分析しよう

(Note)
Y_{it} ：$\log(earning_{it})$
D_{it} ：$schooling_{it}$
u_i ：$intelligence_{it}$

そのまま素直にPooled OLSで分析すると、うまくいかなさそう。

・原因：u_iを無視したこと

だが、u_iが含まれたままでも、それはそれで問題

$\because$ u_iは観測できない

=>なら、「合法的に」u_iに退場してもらえば良い！　＝　Fixed Effects / Random Effects model の導入

Panel Data Analysis 1: Fixed Effects

Fixed Effects model（FE model）

●固定効果モデル（Fixed Effects Model）

- unobserved individual effects(観測不可能な個体固有効果)を、「時間平均をとって引く」ことで取り除き、推定を行う手法

Fixed Effects model（FE model）

(Note)
Y_{it} ：$\log(earning_{it})$
D_{it} ：$schooling_{it}$
u_i ：$intelligence_{it}$

$$Y_{it} = D_{it}\delta_i + u_i + \varepsilon_{it}$$

この回帰式においてu_iは（時間に関して一定であることから）「固定された効果(fixed effect)」と見なすことが出来る。

→　両辺の時間平均をとると、$\sum_{t=1}^{T} u_i = Tu_i$より、$u_i$だけが同じ値のまま残る！

＜定式化＞

$$Y_{it} = D_{it}\delta_i + u_i + \varepsilon_{it} \quad \ldots(1)$$
$$\overline{Y_i} = \overline{D_i}\delta_i + \overline{u_i} + \overline{\varepsilon_i} \quad \ldots(2)$$

$(1) - (2)$より、

$$Y_{it} - \overline{Y_i} = (D_{it} - \overline{D_i})\delta_i + \underbrace{(u_i - \overline{u_i})}_{=0} + (\varepsilon_{it} - \overline{\varepsilon_i})$$

$\because u_i$： 定数

$$\therefore \ddot{Y}_{it} = \ddot{D}_{it}\delta_i + \ddot{\varepsilon}_{it}$$

ただし、$\ddot{X}_{it} = X_{it} - \overline{X_i}$とする。

この回帰式を以て推定を行う

Fixed Effects model (FE model)

FE model の推定量って、本当に問題ない？
→　以下の条件を満たしていれば、$\widehat{\delta_{FE}}$は「一致不偏推定量」になってくれる！

1）「強外生性」の仮定が成立

$$E[\varepsilon_{it}|D_i, u_i] = 0 \quad for\,^\forall t \in \{1,2,\ldots,T\} \qquad (D_i \equiv (D_{i1}, D_{i2}, \ldots, D_{iT}))$$

2）　D_{it}は時間に関して一定の項ではない　<=　当たり前なので、気にする必要は特になし。

$$rank\left(\sum_{t=1}^{T} E\left[\ddot{D_{it}}'\ddot{D_{it}}\right]\right) = K \quad \text{(full rank)}$$

（一応、数式的にはD_{it}が正則行列であることと同値。）

(※)一致推定量：　$\underset{N\to\infty}{\mathrm{plim}}\,\hat{\delta} = \delta$

(※)不偏推定量：　$E\left[\widehat{\sigma^2}\right] = \sigma^2,\ \ \sigma^2 = Var(\varepsilon|D)$

Fixed Effects model (FE model)

FE model の利点：

$Corr(D_i, u_i) \neq 0$、つまり、D_iとu_iの間に相関があっても、欠落変数バイアスを回避して推定が可能になる！

では、FE model の弱点は？

1) Simultaneity (or reverse causality) や 「時間経過で変化するindividual specific effects」には対処がきかない

2)非観測誤差(heterogeneity)u_iが「時間に関して一定ではない」時はバイアスが生じる

3)どうしても係数を推定できない場合がある

Fixed Effects model (FE model)

弱点1)Simultaneity (or reverse causality) や 「時間経過で変化するindividual specific effects」には対処がきかない

e.g.,) Cornwell & Trumbull (1994) “Estimating the Economic Model of Crime with Panel Data.”

・North Carolina州の郡ごとのパネルデータを用いて、人口当たりの警察組織の規模や犯罪の検挙率といった要因が、犯罪の発生率に与える影響を分析した研究。

$$R_{it} = X_{it}\beta + P_{it}\gamma + \alpha_i + \varepsilon_{it}$$

（※）以下、すべて行列。

R_{it}：i郡におけるt期のcrime rate
X_{it}：i郡におけるt期のrelative return to legal activities & P_{it}以外でR_{it}に関係しそうな要因(observable)
P_{it}：i郡におけるt期の検挙や有罪、投獄確率や、刑罰の重さなど (observable)
α_i：i郡における郡固有効果(unobservable)、時間に関して一定

Fixed Effects model (FE model)

弱点1)

$$R_{it} = X_{it}\beta + P_{it}\gamma + \alpha_i + \varepsilon_{it}$$

これを、「個体効果の影響が無い(cross-sectional)」場合、「個体効果がある」場合,の指標にした上で、回帰分析を行った。

$$\overline{R_i} = \overline{X_i}\beta + \overline{P_i}\gamma + \alpha_i + \overline{\varepsilon_i} \quad \ldots \quad (between\ estimator)$$

$$\ddot{R_{it}} = \ddot{X_{it}}\beta + \ddot{P_{it}}\gamma + \ddot{\varepsilon_{it}} \quad \ldots \quad (within\ estimator)$$

$$\left(\because \overline{R_i} = \frac{1}{T}\sum_t R_{it}, \qquad \ddot{R_{it}} = R_{it} - \overline{R_i} \right)$$

以下、説明の簡単のために、$X_{it} = Police_{it}$, $P_{it} = P_{A_{it}}$であるものとする

$Police_{it}$： i郡におけるt期の警察の規模

$P_{A_{it}}$ ： i郡におけるt期の犯罪の検挙率

(Note)

R_{it}：i郡におけるt期のcrime rate

X_{it}：i郡におけるt期のrelative return to legal activities & P_{it}以外でR_{it}に関係しそうな要因(observable)

P_{it}：i郡におけるt期の検挙や有罪、投獄確率や、刑罰の重さなど (observable)

α_i：i郡における郡固有効果(unobservable)、時間に関して一定

Fixed Effects model (FE model)

$$\overline{R_i} = \overline{Police_i}\beta + \overline{P_{A_i}}\gamma + \alpha_i + \overline{\varepsilon_i} \quad \ldots \quad (between\ estimator)$$
$$\ddot{R_{it}} = \ddot{Police_{it}}\beta + \ddot{P_{A_{it}}}\gamma + \ddot{\varepsilon_{it}} \quad \ldots \quad (within\ estimator)$$

between estimator をOLSで分析したものが、Pooled OLS（に相当）、within estimator を OLS で分析したものがFixed Effects model である。

なお、$H_0\colon E[P|X] = 0$が検定により棄却されるため、本分析においては「固定効果の影響あり」とするのが妥当とされている。

(Note)
R_{it}：i郡におけるt期のcrime rate
$Police_{it}$： i郡におけるt期の警察の規模
$P_{A_{it}}$　：i郡におけるt期の犯罪の検挙率
α_i：i郡における郡固有効果(unobservable)、時間に関して一定
$\overline{R_i} = \frac{1}{T}\sum_t R_{it}$, $\ddot{R_{it}} = R_{it} - \overline{R_i}$

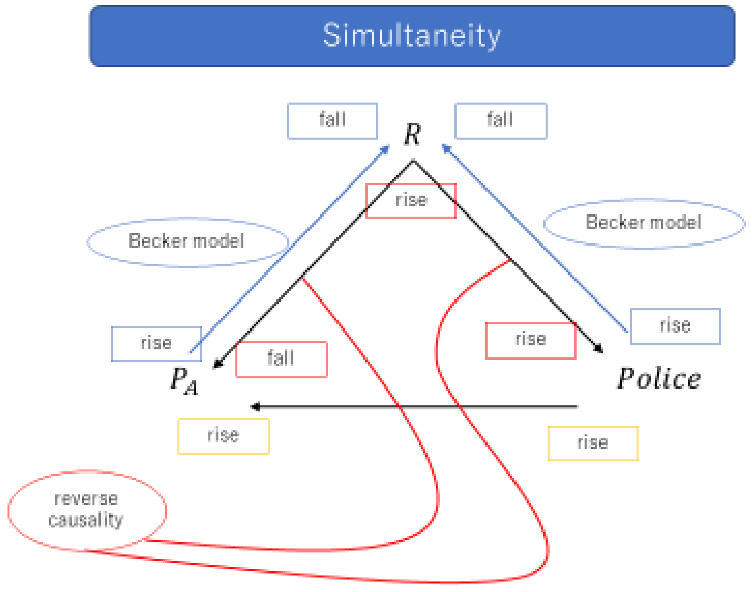
Fixed Effects model (FE model)

弱点1)

$$\overline{R_i} = \overline{Police_i}\beta + \overline{P_{A_i}}\gamma + \alpha_i + \overline{\varepsilon_i} \quad \ldots \quad (between\ estimator)$$
$$\ddot{R_{it}} = \ddot{Police_{it}}\beta + \ddot{P_{A_{it}}}\gamma + \ddot{\varepsilon_{it}} \quad \ldots \quad (within\ estimator)$$

ここで、$R, Police, P_A$の関係に注目。

=>Simultaneity の問題が発生
=>内生性の問題($E[\varepsilon|Police, P_A] \neq 0$)
(↑Hausman 検定でも、$H_0: E \neq 0$が棄却できなかった)
=>2SLS(二段階最小二乗法)でも分析



(Note)
R_{it}：i郡におけるt期のcrime rate
$Police_{it}$： i郡におけるt期の警察の規模
$P_{A_{it}}$　： i郡におけるt期の犯罪の検挙率
α_i：i郡における郡固有効果(unobservable)、時間に関して一定
$\overline{R_i} = \frac{1}{T}\sum_t R_{it}$， $\ddot{R_{it}} = R_{it} - \overline{R_i}$

(! Simultaneity と Reverse Causality の違い)
・Simultaneity(同時性)　$X \leftrightarrow Y$

・Reverse Causality(逆因果)　$Y \to X$　$(but, expected\ as\ X \to Y)$

Fixed Effects model (FE model)

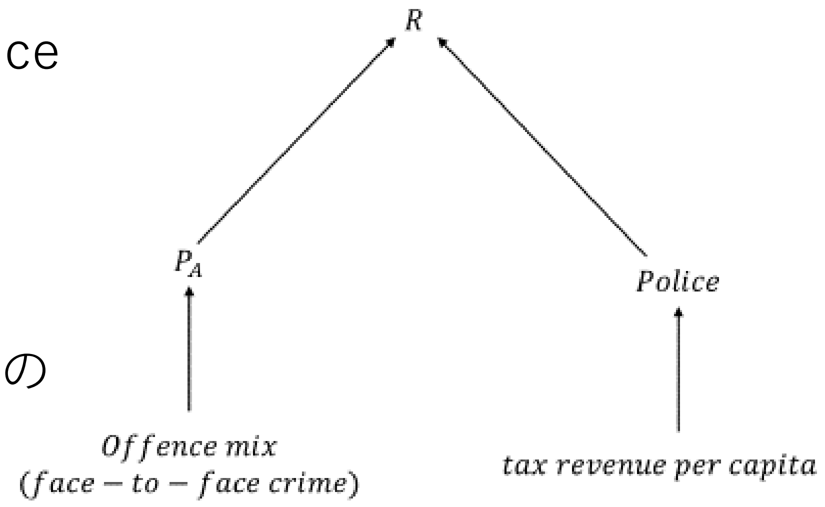
弱点1)

$$\overline{R_i} = \overline{Police_i}\beta + \overline{P_{A_i}}\gamma + \alpha_i + \overline{\varepsilon_i} \quad \dots \quad (between\ estimator)$$

$$\ddot{R}_{it} = \ddot{Police}_{it}\beta + \ddot{P}_{A_{it}}\gamma + \ddot{\varepsilon}_{it} \quad \dots \quad (within\ estimator)$$

操作変数IVとして、それぞれ「Offence mix(強盗など、face-to-face で起こった犯罪)」、
「tax revenue per capita」を利用。

=>普通の固定効果モデルなどでは解決出来ていなかった、同時性の問題の解決を図る



(Note)
R_{it} ：i郡におけるt期のcrime rate
$Police_{it}$ ： i郡におけるt期の警察の規模
$P_{A_{it}}$ ： i郡におけるt期の犯罪の検挙率
α_i ：i郡における郡固有効果(unobservable)、時間に関して一定
$\overline{R_i} = \frac{1}{T}\sum_t R_{it}$, $\ddot{R}_{it} = R_{it} - \overline{R_i}$

Fixed Effects model (FE model)

“no longer statistically significant”

TABLE 3.—RESULTS FROM ESTIMATION
(standard errors in parentheses)

	Between	Within	2SLS (fixed effects)	2SLS (no fixed effects)
CONSTANT	−2.097			[illegible].719
	(2.822)			(8.189)
P_A	−0.648	−0.355	−0.455	−0.507
	(0.088)	(0.032)	(0.618)	[illegible]
P_C	−0.528	−0.282	−0.336	[illegible]
	(0.067)	(0.021)	(0.371)	(0.110)
P_P	0.297	−0.173	−0.196	0.200
	(0.231)	(0.032)	(0.200)	(0.343)
S	−0.236	−0.00245	−0.0298	−0.218
	(0.174)	(0.02612)	(0.0300)	(0.185)
POLICE	0.364	0.413	0.504	0.419
	(0.060)	(0.027)	(0.617)	(0.218)
DENSITY	0.168	0.414	0.291	0.226
	(0.077)	(0.283)	(0.785)	(0.103)

The scale of police is positively correlated to crime rate …!?
(because of simultaneity)

Table :
Cornwell, Christopher, and William N. Trumbull. 1994. “Estimating the Economic Model of Crime with Panel Data.” *Review of Economics and Statistics* 76 (2): 360–66, Retrieved from: https://www.amherst.edu/media/view/121570/original/CornwellTrumbullCrime%2BElasticities.pdf

結局、逆因果関係や同時性がある時、もしくは説明変数と相関があり尚且つ時間変化するような個体固有効果が含まれているような時には、Fixed Effects modelは上手く適用し難い。

Fixed Effects model (FE model)

弱点2)非観測誤差(heterogeneity)u_iが「時間に関して一定ではない」時はバイアスが生じる

$$Y_{it} = D_{it}\delta_i + u_{it} + \varepsilon_{it} \quad \ldots(1)$$
$$\overline{Y_i} = \overline{D_i}\delta_i + \overline{u_i} + \overline{\varepsilon_i} \quad \ldots(2)$$

$(1)-(2)$より、

$$Y_{it} - \overline{Y_i} = (D_{it} - \overline{D_i})\delta_i + (u_{it} - \overline{u_i}) + (\varepsilon_{it} - \overline{\varepsilon_i})$$
$$\ddot{Y_{it}} = \ddot{D_{it}}\delta_i + \ddot{u_{it}} + \ddot{\varepsilon_{it}}$$

u_{it}が観測できない以上、$\ddot{u_{it}}$も分からない。→　仮に$\ddot{\eta_{it}} = \ddot{u_{it}} + \ddot{\varepsilon_{it}}$とまとめて回帰をしようものなら、欠落変数バイアス

$$\frac{Covariance(\ddot{D_{it}}, \ddot{u_{it}})}{Var(\ddot{D_{it}})}$$

が生じてしまう

Fixed Effects model (FE model)

弱点3)どうしても係数を推定できない場合がある

・以下のようなタイプの項に関しては、FE modelでは係数推定が出来ない

・時間変化をしない項

e.g.,) u_i, 定数項, 属性ダミー(i.e., 性別, 人種etc.)

属性ダミーについては、時点時点ダミー(i.e., 1994年ダミー)との交差項をとることで係数の推定は可能になる。ただし、解釈の仕方には注意。

・時間変化の幅が一定な項

e.g.,) 年齢, 経験年数（1年経過毎に、「すべての標本に対して」+1ずつ増加）

すべての年度の時間ダミーを含むようなモデルを考えるときには、このタイプの項は推定が出来なくなる。

Panel Data Analysis 2 : Random Effects

Random Effects Model

●変量効果モデル（**Random Effect Model**）

- 観測されない個体固有効果を独立変数とは独立としてランダムに発生しているものとみなして推計を行う。

→ 個体特性を誤差項$\boldsymbol{u_i}$として扱う

Random Effects model（RE model）

$$Y_{it} = D_{it}\delta_i + u_i + \varepsilon_{it}$$

今度はu_iを「確率変数」と見なして、定式化を図る。

→　すべてのu_iに共通の平均値μ_iを中心に分散$\sigma_{u_i}^2$で分布する確率変数、と見る。

＜定式化＞（詳細は割愛。）

$$Y_{it} = D_{it}\delta_i + (\mu_i + \alpha) + \varepsilon_{it}$$

$$\therefore Y_{it} = D_{it}\delta_i + \alpha + (\mu_i + \varepsilon_{it})$$

ここから変数変換を施して、最終的に以下を得る。

$$\boxed{Y_{it}^* = D_{it}^*\delta_i + (1-\theta) + \varepsilon_{it}}$$

$$\left(Y_{it}^* = Y_{it} - \theta\overline{Y}_i, X_{it}^* = X_{it} - \theta\overline{X}_i, \qquad \theta = 1 - \sqrt{\frac{\sigma_\varepsilon^2}{T\sigma_u^2 + \sigma_\varepsilon^2}} \right)$$

メモ：$\mu_i + \varepsilon_{it}$は合成誤差(a composite error term)と呼ばれる

メモ：θの推定が必要になる。ガウス・マルコフの仮定を満たすように変数変換を施してからOLSを適用する手法のことをGeneralized Least Square (GLS, 一般化最小二乗法)と言い、また今回のような推定の過程を含むGLSをFeasible GLS (FGLS)と呼ぶ。

Random Effects model (RE model)

RE model の利点：

FE model では推定できなかった、定数項や時間に関して一定な変数の係数などを推定できる。

RE model を適用するための条件：

$$Corr(D_{it}, u_i) = 0$$

要は、D_i, u_iの間には相関が無い、ということが条件。

RE 推定量が一致不偏推定量であるための条件：

・強外生性

$$E[\varepsilon_{it} | D_i, u_i] = 0$$

Panel Data Analysis: which one to use?

分析手法の使い分け

$$Y_{it} = D_{it}\delta_i + u_i + \varepsilon_{it}$$

・もし個体特有の効果（ u_i ）が無視できる程度（ $u_i \approx 0$）のモノだとしたら、Pooled OLSも使える。
・ただ、これをどうやって見分ける？
・ $u_i \neq 0$として、Fixed Effects と Random Effects をどうやって使い分ける？

→検定の利用

分析手法の使い分け

$$Y_{it} = D_{it}\delta_i + u_i + \varepsilon_{it}$$

1) そもそもPooled OLSをつかって良さそうか？

F検定（Breusch-Pagan検定）で、個体個別効果の有無を確認

H_0：「個別効果なし」　H_1：「個別効果あり」

H_0棄却 → Fixed Effects / Random Effects , H_0を棄却できない → Pooled OLS

2) Fixed Effects と Random Effects, どちらを使うか？

Hausman 検定で、D_iとu_iの相関の有無を確認

H_0：「$Corr(D_i, u_i) = 0$」　H_1：「$Corr(D_i, u_i) \neq 0$」

H_0棄却 → Fixed Effects,　H_0を棄却できない → Random Effects

あくまで、一つの指標として参考に。

Panel Data Analysis -in practice-

Panel data analysis _ example

- Cornwell & Rupert (1997) “Unobservable Individual Effects, Marriage and the Earnings of Young Men”

・男性に関して、婚姻が賃金に与える影響を分析。「結婚が生産性（賃金）を向上させる」は本当か？

$$y_{it} = \delta + X_{it}\beta + M_{it}\gamma + \alpha_i + \varepsilon_{it}$$

（※）以下、すべて行列。
y_{it}：t期におけるiの賃金（log）
X_{it}： t期におけるiの諸々の説明変数
M_{it}： t期におけるiの婚姻状況（ダミー変数）
α_i： t期におけるiに固有な効果（unobserved）

イメージ：

$$\log(wage_{it}) = \delta + experience_{it}\beta_1 + \cdots + education_{it}\beta_k + Married_{it}\gamma_1 + \cdots + \alpha_i + \varepsilon_{it}$$

Panel data analysis _ example

y_{it}：t期におけるiの賃金（log）
X_{it}： t期におけるiの諸々の説明変数
M_{it}： t期におけるiの婚姻状況（ダミー変数）
α_i： t期におけるiに固有な効果（unobserved）

$$y_{it} = \delta + X_{it}\beta + M_{it}\gamma + \alpha_i + \varepsilon_{it}$$

Random Effects model(FGLS) vs. Fixed Effects model

$$Corr(X_i, \alpha_i) = 0 \quad or \quad Corr(X_i, \alpha_i) \neq 0 \ ?$$

-> Wu-Hausman 検定

今回は$\chi^2 = 111.9$となっており、帰無仮説
H_0: $Corr(X_i, \alpha_i) = 0$
が棄却される。
→ Fixed Effects model の方が適切と判断

TABLE II
Estimated Wage Regressions
(Standard Errors in Parentheses)

Variable	(1) FGLS	(2) Within	(3) Within	(4) Within
Married	0.083 (0.022)	0.056 (0.026)	0.051 (0.026)	0.033 (0.028)
Divorced	0.064 (0.033)	0.062 (0.036)	0.057 (0.036)	0.040 (0.038)
Years Married				–0.005 (0.006)
Years Married^2				–0.0003 (0.0003)
Years Divorced				–0.014 (0.008)
Experience	0.027 (0.004)	0.027 (0.004)	0.024 (0.004)	0.021 (0.005)
Experience^2	–0.001 (0.0001)	–0.001 (0.0002)	–0.001 (0.0002)	–0.001 (0.0002)
Tenure			0.013 (0.004)	0.011 (0.004)
Tenure^2			–0.0006 (0.0002)	–0.0005 (0.0002)
South	–0.091 (0.019)	–0.121 (0.034)	–0.117 (0.034)	–0.118 (0.034)
Urban	0.137 (0.017)	0.057 (0.024)	0.059 (0.024)	0.059 (0.024)
Union	0.109 (0.015)	0.106 (0.018)	0.102 (0.018)	0.103 (0.018)
Dependents	0.052 (0.017)	0.052 (0.019)	0.048 (0.019)	0.047 (0.020)
No High School	–0.325 (0.057)			
Some High School	–0.148 (0.032)			
Some College	0.091 (0.028)			
College Grad	0.278 (0.034)			
Post-College	0.322 (0.041)			
Standard error	[illegible]	0.212	0.212	0.211
χ^2_{27}	111.9			

Cornwell & Rupert.(1997). “Unobserved individual effects, marriage and the earnings of young men”. Economic inquiry. 1997, vol. 35, no. 2, p. 285–294.

Panel data analysis _ example

y_{it} ：t期におけるiの賃金（log）
X_{it} ： t期におけるiの諸々の説明変数
M_{it} ： t期におけるiの婚姻状況（ダミー変数）
α_i ： t期におけるiに固有な効果（unobserved）

$$y_{it} = \delta + X_{it}\beta + M_{it}\gamma + \alpha_i + \varepsilon_{it}$$

Fixed Effects model：

$$\ddot{y_{it}} = \ddot{X_{it}}\beta^* + \ddot{M_{it}}\gamma + \ddot{\varepsilon_{it}} \quad (within\ estimator)$$

! 時間に関して一定であった変数は within estimator では消えている
e.g.,) education

著者はFixed Effects model での分析を
1) そのまま　2) 変数$Tenure, Tenure^2$を追加　3) 変数$YearsMarried, YearsMarried^2, YearsDivorced$を追加
の3通り行っている。

(1) Random Effects model での分析
(2) Fixed Effects model での分析その１
(3) Fixed Effects model での分析（Tenureを追加）
(4) Fixed Effects model での分析（Year Marriedなど追加）

TABLE II
Estimated Wage Regressions
(Standard Errors in Parentheses)

Variable	(1) FGLS	(2) Within	(3) Within	(4) Within
Married	0.083 (0.022)	0.056 (0.026)	0.051 (0.026)	0.033 (0.028)
Divorced	0.064 (0.033)	0.062 (0.036)	0.057 (0.036)	0.040 (0.038)
Years Married				–0.005 (0.006)
Years Married^2				–0.0003 (0.0003)
Years Divorced				–0.014 (0.008)
Experience	0.027 (0.004)	0.027 (0.004)	0.024 (0.004)	0.021 (0.005)
Experience^2	–0.001 (0.0001)	–0.001 (0.0002)	–0.001 (0.0002)	–0.001 (0.0002)
Tenure			0.013 (0.004)	0.011 (0.004)
Tenure^2			–0.0006 (0.0002)	–0.0005 (0.0002)
South	–0.091 (0.019)	–0.121 (0.034)	–0.117 (0.034)	–0.118 (0.034)
Urban	0.137 (0.017)	0.057 (0.024)	0.059 (0.024)	0.059 (0.024)
Union	0.109 (0.015)	0.106 (0.018)	0.102 (0.018)	0.103 (0.018)
Dependents	0.052 (0.017)	0.052 (0.019)	0.048 (0.019)	0.047 (0.020)
No High School	–0.325 (0.057)			
Some High School	–0.148 (0.032)			
Some College	0.091 (0.028)			
College Grad	0.278 (0.034)			
Post-College	0.322 (0.041)			
Standard error	0.215	0.212	0.212	0.211
χ^2_{27}	111.9			

Cornwell & Rupert.(1997). “Unobserved individual effects, marriage and the earnings of young men”. Economic inquiry. 1997, vol. 35, no. 2, p. 285–294.

Panel data analysis _ example

結果：

$Married$の係数…FGLS >> FE(2)>FE(3)>FE(4)

△　結婚が生産性（賃金）を向上させる
〇　賃金上昇は、婚姻関係と賃金、双方に相関のある、観測されない個体固有効果の影響

この時点で、5%水準で、統計的に有意ではなくなっている

Variable	(1) FGLS	(2) Within	(3) Within	(4) Within
Married	0.083 (0.022)	0.056 (0.026)	0.051 (0.026)	0.033 (0.028)

Cornwell & Rupert.(1997). “Unobserved individual effects, marriage and the earnings of young men”. Economic inquiry. 1997, vol. 35, no. 2, p. 285–294.

Appendix

Appendix 1: 回帰式の行列形と、$\boldsymbol{X_i}$の影響

$$Y_{it} = D_{it}\delta_i + X_i\gamma_i + u_i + \varepsilon_{it}$$

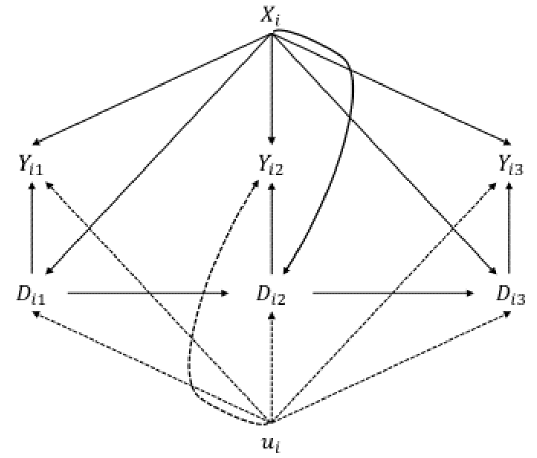
$$Y_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{it} \\ \vdots \\ Y_{iT} \end{pmatrix}_{T\times 1},\ D_{it} = \begin{pmatrix} D_{it,1} & \cdots & D_{it,j} & \cdots & D_{it,K} \end{pmatrix}_{1\times K},\ D_i = \begin{pmatrix} D_{i1} \\ \vdots \\ D_{it} \\ \vdots \\ D_{iT} \end{pmatrix}_{T\times K},\ \delta_i = \begin{pmatrix} \delta_{i1} \\ \vdots \\ \delta_{ij} \\ \vdots \\ \delta_{iK} \end{pmatrix}_{K\times 1},\ \varepsilon_i = \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{it} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}_{T\times 1},$$

X_i, γ_i, u_iは行列ではない普通の項

cross-sectional な見方でこれを分析すると、

$$Y_{it} = D_{it}\delta_{it} + \eta_{it}, \qquad \eta_{it} = X_i\gamma_i + u_i + \varepsilon_{it}$$

$$\widehat{\delta_{it}} = \delta_{it} + \frac{Covariance(D_{it}, X_i\gamma_i + u_i)}{Var(D_{it})}$$



DAGより、 $Covariance(D_{it}, X_i\gamma_i + u_i) \neq 0$ なので、欠落変数バイアスはかかります。

Appendix 1: 回帰式の行列形と$\boldsymbol{X_i}$の影響(続)

Fixed Effects model を適用すると、t に依らず一定の項 $X_i\gamma_i + u_i$ は相変わらず消えるので、

$$\ddot{Y_{it}} = \ddot{D_{it}}\delta_i + \ddot{\varepsilon_{it}}$$

後の議論は同じものが適用できます。

Appendix 2: FE推定量=一致不偏推定量

Condition

1）個体固有誤差(idiosyncratic error) ε_{it} と独立変数 D_i, 非観測誤差 u_i は独立

「強外生性(strict exogeneity)」

$$E[\varepsilon_{it}|D_i, u_i] = 0 \quad for\ t = 1,2,\dots,T \quad D_i \equiv (D_{i1}, D_{i2}, \dots, D_{iT})$$

2) 説明変数 $D_i \equiv (D_{i1}, D_{i2}, \dots, D_{iT})$ は、すべて時間変化するD_{it}で構成されている

$$rank \sum_{t=1}^{T} E[\,{}^t\ddot{D}_{it}\ddot{D}_{it}] = K \quad (\dots \text{full rank})$$

<Prf.>

もし、時間変化しない物が入っていると、$K \times K$行列 ${}^t\ddot{D}_{it}\ddot{D}_{it}$のどこかの列、ないし行に0のみで構成される物が、全てのtに関して現れる。（しかも、全てのtについて同じ行、列に出てくる。）なので、tに関して足し合わせても、どうしても全ての要素が0の行、または列が出てくる。フルランクにならない。（実際、D_{it}が正則であることと同値。））

Appendix 3: Fixed Effects model - $\hat{\delta}$の計算

$$Y_{it} = D_{it}\delta_i + +u_i + \varepsilon_{it}$$

$$Y_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{it} \\ \vdots \\ Y_{iT} \end{pmatrix}_{T\times 1}, D_{it} = \begin{pmatrix} D_{it,1} & \cdots & D_{it,j} & \cdots & D_{it,K} \end{pmatrix}_{1\times K}, \quad D_i = \begin{pmatrix} D_{i1} \\ \vdots \\ D_{it} \\ \vdots \\ D_{iT} \end{pmatrix}_{T\times K},$$

$$\delta_i = \begin{pmatrix} \delta_{i1} \\ \vdots \\ \delta_{ij} \\ \vdots \\ \delta_{iK} \end{pmatrix}_{K\times 1}, \quad \varepsilon_i = \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{it} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}_{T\times 1}$$

$$\left(\hat{\delta}, \widehat{u_1}, \dots \widehat{u_N}\right) = \arg \min_{\beta, m_1, \dots m_N} \sum_{i=1}^{N} \sum_{t=1}^{T} (Y_{it} - D_{it}\beta - m_i)^2$$

これを満たす$\hat{\delta}$を求める。（別途、PDFを掲載。）