# 恐竜本 補足資料

前川 大空 *

2025 年 10 月 15 日

## 目次



* 一橋大学経済学部 4 年, 五年一貫専修コース公共経済プログラム

## はしがき

**■p.i 構造推定アプローチ**　経済理論による均衡条件の構造型の導出 → その誘導型の導出 → 最小 2 乗推定量 → 操作変数法 → 一般化積率法 (GMM) と進む流れ (Ex. Wooldridge, 2010). 潜在結果アプローチはこの中に埋め込まれる型で説明されている, らしい. DDR [*1]にせよ『構造推定アプローチ』への批判が最近の潮流なのかもしれない.

**■p.i 構造推定アプローチの問題点**　既存教科書の説明スタイルは, 制度の記述から始まる 近年のミクロ経済学の実証研究における因果推論の実践とは必ずしも一致する型とはなっていない. 一般均衡理論のみならない分析, ゲームなどを想定する点で異なるといったはなし? → 実制度の話を意図しているだけ?

**■p.ii 『経済学訛り』の因果推論**　経済学においては DiD や RDD など, 疑似実験の手法が頻繁に活用される.

**■p.iii モチベーション**　『既存の知識は暗黙知も含めてドキュメント化してさっさと共有してしまおう. 人々の貴重な時間はもっと新しくて未解決の問題について考えるために使おう』. マジでアツい.

**■p.iv コンテンツについて**　R の実装についてはウェブ付録の **実践編** において記載されている. **Lec5,6** の実装でも参考に出来るかもしれない.

## 序章 経済学の因果推論アプローチ

**■p.1 教科書の流れ**　第 2 ～ 5 章: RCT, 第 6 ～ 8 章: DiD, 第 9 ～ 11 章: RDD
第 1 章: 潜在結果モデル導入, 第 2 章: ATE の推定, 第 3 章: 統計的推測, 第 4 章: 非遵守者, 第 5 章: 実践

**■p.3 潜在結果モデルの導入法**　『筆者らは, 本書のように潜在結果モデルだけに依拠して議論を進めたほうが初学者にとっては理解しやすいのではないかと考えている.』 構造型を背後に (明示的には) 想定しない, DDR でも同じような議論は見受けられた.

**■p.4 『経済学訛り』の因果推論**　行動反応を考えることが訛りの原因, 非遵守者を考えるのはこのため.

**■p.4 同値観測性の問題**　Ex. 上向きの需要関数, **Lec1** のひとまとめにしたことによる上向き傾向. 解決策として識別のための誘導型が導入された.

**■p.5 信頼性革命**　自然実験や疑似実験に注目すべき, との近年の計量経済学研究における潮流. 構造推定アプローチはここに対応できない点でも近年の評価を落としている?

**■p.5 提示されている問い**　本書を読み進めるにあたって答えを出せるようになろう.

- RDD を局所線型モデルで推定するとき, 推定対象範囲を定めるためのバンド幅はの選び方は?
- 処置群と統制群を分けるスコア変数が離散的なときはどのように対処すべきだろうか?
- DiD で因果効果を推定しようとするとき, 処置が個体間で一斉に行われるのではなくタイミングがずれている場合にはどのような定式化を用いるべきだろうか?

[*1] 末石 (2024), データ駆動型回帰分析, 日本評論社, 第 1 版

■p.7 **構造推定アプローチと IV(GMM) の別**　構造型を内生変数について解き, 外生変数だけの誘導型に直して推定するのが構造推定アプローチ. 一方で, ショックについて解き, そのモーメントに関する条件 (積率条件) を利用してパラメータ推定するのが GMM, 定式化の範囲を限定したものが IV method.

■p.8 **構造推定アプローチと潜在結果アプローチの別**　構造推定モデルは推定対象である誘導型を導出するために, 経済モデルから導き出された構造型を利用している時点で, 検証不可能な仮定を置いていることとなる. 一方の潜在結果アプローチは実験デザインなどの制約を変更することによってのみ推定の方法を変更しており, 経済理論は結果の解釈にのみ用いられる.

■p.9 **『誘導型の推定』との言い方の不適切性**　潜在結果モデルは経済理論 (構造型) から導出されるとは限らないため, 誤解を招く不適切ないいかた.

# 1　潜在結果モデルと因果効果

■p.12 **諸定義の導入**　以下のように定義されている.

**Def. 実験, 無作為化実験**

個体への処置割当が既知な状況のこと.
特に個体への処置割当が無作為に行われているようなことが既知な状況を無作為化実験と呼ぶ.

**Def. タイプ**

潜在結果はあらかじめ個体ごとに定まっているものとして, この対応結果を潜在結果のタイプと呼ぶ.

**Def. 因果関係**

タイプの集合のうち幾つかのタイプに属するものを因果効果として定義する.

■p.16 **潜在結果と健在結果の別**　潜在結果は関数だが, 健在結果は値である, $\mathcal{Z}$ について足したため.

■p.17 **潜在結果モデル**　$(\mathbf{Y}, \mathbf{Z}, \mathbf{W}, \mathbf{Y}^*(\cdot))$ を潜在結果モデルと呼ぶ. うち潜在結果 $\mathbf{Y}^*(\cdot)$ は観測不可能.

■p.18 SUTVA　2 項処置モデルとして定義してしまおう:

**Assumption: SUTVA**

潜在的結果が以下のように表されることを指す.

$$Y_i = Y_i^*(Z_i) = \begin{cases} Y_i^*(1) & \text{if } Z_i = 1 \\ Y_i^*(0) & \text{if } Z_i = 0 \end{cases} \quad \text{where } Z_i \in \{0, 1\}$$

つまり, 以下の 2 条件がこの式で表現されている:

1. no spillover: 潜在的結果は他者の処置の影響を受けない (引数が $Z_i$ のみ)
2. consistency: 処置が施されたとき, 観測されるのは潜在結果をその処置で評価した値
3. binary treatment: 処置が均一 (デジタルな処置)

**■p.21 処置割当メカニズム**　処置割当メカニズムとは, $(\mathbf{W}, \mathbf{Y}^*(0), \mathbf{Y}^*(1))$ から $\mathbf{Z}$ への行の交換に対して不変な非負の関数 $\mathbb{P}[\mathbf{Z} \mid \mathbf{W}, \mathbf{Y}^*(0), \mathbf{Y}^*(1)]$ で, 任意の $(\mathbf{W}, \mathbf{Y}^*(0), \mathbf{Y}^*(1))$ に対して以下が成り立つものである.

$$\sum_{\mathbf{Z} \in \{0,1\}^n} \mathbb{P}[\mathbf{Z} \mid \mathbf{W}, \mathbf{Y}^*(0), \mathbf{Y}^*(1)] = 1 \tag{1.3}$$

つまり, プロファイルの選び取り方全体にわたって定義された確率測度と言えよう. $\mathbf{Y}^*(0), \mathbf{Y}^*(1)$ に依存しており, これらは正規処置割当メカニズムを考慮すれば除外されることとなる.

**■p.22 個体処置割当確率**　$p_i$. $2^{n-1}$ 個を足し合わせており $1 - p_i$ が個人非割り当て確率となる.

**■p.22 傾向スコア**　$e(\mathbf{w})$. 引数として顕わには $\mathbf{w}$ のみがあるが, 実際には $\mathbf{Y}^*(0), \mathbf{Y}^*(1)$ にも依存する.

**■p.23 正規処置割当メカニズム**　以下の条件を満たすものが正規処置割当メカニズムと呼ばれる.

1. 個別割当: $p_i$ が $p_i[\mathbf{W}, \mathbf{Y}^*(1), \mathbf{Y}^*(0)] = q[\mathbf{W}_i, \mathbf{Y}_i^*(1), \mathbf{Y}_i^*(0)]$ と個人の情報のみに依存する
2. 確率割当: すべての個人の $p_i$ が $0 < p_i[\mathbf{W}, \mathbf{Y}^*(1), \mathbf{Y}^*(0)] < 1$
3. 条件付き独立割当: $p_i$ が 潜在結果に依存せず, $\mathbb{P}[\mathbf{Z} \mid \mathbf{W}, \mathbf{Y}^*(0), \mathbf{Y}^*(1)] = \mathbb{P}[\mathbf{Z} \mid \mathbf{W}]$

これは以下のように単純化された:

**正規処置割当メカニズム**

ある集合 $\mathcal{A}$ が存在し, 任意の $(\mathbf{Z}, \mathbf{W}, \mathbf{Y}^*(0), \mathbf{Y}^*(1)) \in \mathcal{A}$ について.

$$\mathbb{P}[\mathbf{Z} \mid \mathbf{W}, \mathbf{Y}^*(0), \mathbf{Y}^*(1)] \propto \prod_{i=1}^{n} q(\mathbf{W}_i)^{Z_i} [1 - q(\mathbf{W}_i)]^{1-Z_i} \tag{1.6}$$

であり, それ以外の処置割当メカニズムの値は 0.

このとき, 傾向スコアはその個体の個体処置割当確率と一致する: $e(\mathbf{w}) = q(\mathbf{w})$.

**■p.24 抽出・割当に起因する不確実性**　母集団から標本 $n$ を抽出する際の不確実性が **抽出に起因する不確実性** で, 処置 $Z_i$ 決定に際する不確実性が **割当に起因する不確実性** である. **p.25** の説明に詳しい.

**■p.27 識別**　2 項処置モデルを前提とした識別の具体的説明がかなりのページを割いて行われる.

**大まかな認識: 識別可能性**

観測されるデータの同時分布が既知の時, $\boldsymbol{\theta}$ の値が一意に定まるならば, $\boldsymbol{\theta}$ は識別されるという.

Ex. **学習塾に通う 5 人を標本とした ATE の識別**　2 項処置モデルで, 3 人に特別講習を **無作為に** 割り当てることを考える. 処置割当メカニズム ($\mathbf{Z}$ に割り当てられる条件付確率測度) は $\mathbf{P_Z}$ で表される. ありうる $\mathbf{Y}^*, \mathbf{P_Z}$ の集合を $\mathcal{M}$ と書く. ここで $m = (\mathbf{y}^*, \mathbf{p_Z})$ に値が定まると **モデルは一意に定まる**. [*2] モデル, 特に $\mathbf{y}^*$ が一意に定まったことで, **今回識別の対象である ATE も一意に定まる**:

$$\theta = \frac{1}{n} \sum_{i=1}^{n} [y_i^*(1) - y_i^*(0)]$$

[*2] $\mathbf{Z}$ の分布 $\mathbf{p_Z}$ は無作為割当から均等に割り振られ, 常に $\mathbf{p_Z}[\mathbf{Z}] = \dfrac{1}{{}_5C_3}$ と一定の値をとるために, 単に『値』と書かれる.

**しかしこれは観測不可能な $\mathbf{y}^*$ が含まれているため，識別に関しては現状何も言えない**．そこで観測されるデータの同時分布 $\mathbf{p_{ZY}}$ について考えると，これも一意に定まり，$\mathbf{p_{ZY}}[\mathbf{z},\mathbf{y}] = \mathbf{p_Z}[\mathbf{z}]\mathbb{1}\{\mathbf{y}^*(\mathbf{z}) = \mathbf{y}\}$．$\phi$ にはこの (**観測可能な**) 同時分布 $\mathbf{p_{ZY}}$ と観測可能なすべての情報が含まれている．この例では割当が分析者にとって既知であるため，$\mathbf{p_Z}$ も $\phi$ に含まれる．モデルの値 $m = (\mathbf{y}^*, \mathbf{p_Z})$ を定めれば，$\phi$ **も一意に定まる**．

しかし，**複数の異なるモデルが同じ** $\phi$ **を生成する可能性がある**．ここで，構造 $s(\phi,\theta) \subset \mathcal{M}$ を，$\phi,\theta$ を生成する **モデル** $m$ **の値の集合** と定義し，この問題を明確化するために **観測上同値性** を定義する:

**今回の例における定義: 観測上同値性**

ATE $\theta,\widetilde{\theta}$ について，ある $\phi$ が存在して，構造 $s(\phi,\theta), s(\phi,\widetilde{\theta})$ がともに空でないとき，観測上同値という．

つまり，$\theta,\widetilde{\theta}$ を生成する異なるモデル $(m,m')$ が観測上同じ情報 $\phi$ を生成する可能性があり，逆に言えば，その異なる ATE を生成する 2 モデルのうちどちらから $\phi$ が生成されているかを，観測される情報からでは **原理的に** (非常に小さい可能性だろうが) 区別できない場合があることを意味する．観測上同値性を回避するための方法としては以下が挙げられている:

1. 割当メカニズムの知識 $\mathbf{p_Z}$ を得ること: **この例では既に満たしている**
2. モデルの値の幅 ($\mathcal{M}$) を狭くする: 正規処置割当メカニズムへの限定，$\mathbf{Z} \perp\!\!\!\perp \mathbf{Y}^*$ **の条件追加 (今回)**

さて，例に立ち戻り，$\mathbf{p_Z}$ が $\phi$ に含まれていることを思い出そう．定義より $\mathbf{p_{ZY}}$ も $\phi$ に含まれる．$\mathbf{Z} \perp\!\!\!\perp \mathbf{Y}^*$ **を利用することによって，観測可能な変数で構成された**，処置群と統制群の平均値の差は:

$$\mathbb{E}\Big[\frac{1}{n_1}\sum_{i=1}^{n} Z_i Y_i - \frac{1}{n_0}\sum_{i=1}^{n}(1-Z_i)Y_i\Big] = \theta(\mathbf{Y}^*) = \theta \tag{1.7}$$

と **興味の対象である ATE に対して不偏性を持つ** ことが示される．つまり，今回の例では，情報 $\phi$ ($\mathbf{p_{ZY}}, \mathbf{p_Z}, \mathbf{y}$ など) が 1 つ与えられると，それと整合的なモデルの値 $m = (\mathbf{y}^*, \mathbf{p_Z})$ が含意するパラメータ $\theta$ の値は一意に定められる．より詳しく言えば，異なるモデル $(m,m') = ((\mathbf{y}^*, \mathbf{p_Z}), (\mathbf{y}^{*\prime}, \mathbf{p_Z}))$ で，ATE $(\theta,\widetilde{\theta}) = (\theta(\mathbf{y}^*), \widetilde{\theta}(\mathbf{y}^{*\prime}))$ について $\theta \neq \widetilde{\theta}$ となるケースを考えると，情報 $\phi$ ($\mathbf{p_{ZY}}, \mathbf{p_Z}, \mathbf{y}$ など) は，特に割当の独立性による $\mathbf{y}$ の不一致によりそれぞれのモデルで異なり ($\mathbf{y} \neq \mathbf{y}' \implies \phi \neq \phi'$)，従って $(\theta,\widetilde{\theta})$ は観測的同値になり得ない．つまり，$\mathbf{Z} \perp\!\!\!\perp \mathbf{Y}^*$ **は「$\mathbf{y}$ をそのままに $\mathbf{y}^*$ を変える」ことを許してくれない** のである．この状態をもって，ATE はこの例において **識別** されていると定義する．

**今回の例における定義: 識別可能性**

情報を増やす，またはモデルのとりうる値を減らすなどといった何らかの方法を経て，ある ATE $\theta$ に対してそれと観測上同値な別のパラメータ $\widetilde{\theta}$ が存在しないとき，そのパラメータ $\theta$ が識別されるという．

今回は，$\mathbf{p_Z}$ が既知 (**情報を増やす**)，($\mathbf{Z} \perp\!\!\!\perp \mathbf{Y}^*$) であることによって $\mathcal{M}$ の範囲が狭まっており，故に識別が可能となったことを認識せよ．

**定義: $\Theta$ のグローバルな識別可能性**

モデル $\mathcal{M}$ が生成しうるパラメータの集合 $\Theta$ の任意の要素 $\theta$ が識別されること．

**■識別の構成的な証明**　例は，観測可能なものが，興味のあるパラメータに一致することを示すものだった．

**定義: 推定対象 (estimand)**

必ずしも興味はないが, 情報 $\phi$ に含まれる観測から構成できる対象一般のこと.

この例の平均値の差 **(1.7) 式** も推定対象である.

**定義: 識別の構成的な証明**

推定対象の 1 つが仮定のもとで興味のあるパラメータに一致することを示す方法.

つまり, 例の識別の証明手順は, 構成的な証明である.

**■p.32 識別が成立しないケース**　先述の例から, 無差別割当を取り去っている. これによって, 異なる $\theta, \widetilde{\theta}$ を生成する異なるモデル $(m, m') = ((\mathbf{y}^*, \mathbf{p_Z}), (\mathbf{y}^{*\prime}, \mathbf{p_Z}))$ が $\mathbf{p_{ZY^*}}$ の違いをのこして, 観測上同じ情報 $\phi$ ($\mathbf{p_{ZY}}, \mathbf{p_Z}, \mathbf{y}$ など) を生成しており, モデルを区別できないために, 識別が出来ていない.

**■Lewbel(2019): 一般的な識別の定義**　$\mathcal{M}$ が何故か関数の集合として定義されているが意図が分からない.

## 2　無作為化実験

**■p.35 答えたい疑問点**

1. 実験を行うことで ATE を推定できる理由は?
2. 無作為化実験・層化無作為化実験など実験方法における差異の, 推定への影響は?

**■p.36 実験データの定義**　本書では, 「処置割当メカニズムが**確率的かつ既知**」な状況から得られたデータを「(無作為化) 実験データ」と呼び, それ以外の状況で得られたデータを「観察データ」と呼ぶ. つまり, 先述の学習塾の例で得られるデータは実験データといえる.

**■p.36 自然実験, 疑似実験**　本書では以下のような定義をされている. **図 2.1** に詳しい.

1. **自然実験**: 処置割当メカニズムは確率的かつ既知だが, 設計者が分析者以外の主体である状況
2. **疑似実験**: 処置割当メカニズムは**未知**だが, 望ましい性質 (Ex. 条件付独立割当) が正当化できる状況

**■p.38 無作為化実験**　本書では以下のような定義をされている.

1. **無作為化実験**: 確率割当であり, 関数形が分析者にとって既知な処置割当メカニズム
2. **古典的な無作為化実験**: 無作為化実験のうち, **正規処置割当メカニズム** であるもの

関数形については実験の設計段階で操作可能と考えることも可能.

**■p.38 古典的な無作為化実験**　4 つの代表的な例が提示されている.
Ex.1 **完全無作為化実験**　$n$ 個から $n_1$ 個の個体を無作為に抽出して介入を行う実験, 学習塾の例はこれ.
Ex.2 **層化無作為化実験**　$n$ 個の個体をまず共変量の値に従って $G$ 個の層に分け, それぞれの層 $g$ の中から規定の数 $n_{1g}$ の個体を抽出して, 全体としては $n_1$ の個体に対して介入を行う実験
Ex.3 **対応化無作為化実験**　層化無作為化実験の $n_g = 2$, $n_{1g} = 1$ の特殊ケース

Ex.4 **クラスター化無作為化実験** 層化無作為化実験と同様に個体を共変量に基づいてクラスターに分け, 規定数のクラスターを無作為に抽出し, **そのクラスターに含まれる個体すべて** に介入を行う実験

■p.40 **クラスター化無作為化実験の特徴**

1. 実施コストが比較的安価
   人材を派遣すべきエリア数を減らせる
2. クラスター単位の効果測定に利用可能
   Ex. 学校でのワクチン接種: 生徒個人への効果は波及効果で測定困難 (**SUTVA に違反**),
   学校単位の効果測定は可能

■p.41 **明確な帰無仮説** 以下のように定義される.

**定義: 明確な帰無仮説**

その仮説下で, 顕在結果から各個体の観測されていない潜在結果がすべてわかるような帰無仮説

具体的には, 以下のような帰無仮説が当てはまる.

$$H_0: \ Y_i^*(0) = Y_i^*(1), \quad i = 1, \dots, n.$$

ATE が 0, は明確な帰無仮説とはならないことに注意.

■p.42 **フィッシャーの $p$ 値**

■p.59 **完全無作為化実験における確率割当** 全ての処置ベクトルに等しい確率を割り当てる, とあるが, これはあくまで処置群の個体数が $n$ になる処置ベクトルに限定した話であることを忘れないように.

■p.60 **バランスが取れた層を重視するウェイト** 他にして一定で $e_g = 1/2$ で最大.

■p.60 **答えたい疑問** フィッシャーの $p$ 値や ATE の推定量の, 完全無作為化実験との違いは?

■p.61 **層化無作為化実験におけるフィッシャーの $p$ 値** $p = \mathbb{P}[|T_\lambda(\mathbf{Z}, \mathbf{y})| \geq |T_\lambda(\mathbf{z}, \mathbf{y})|]$ で計算される.

■p.61 **層化無作為化実験における ATE の不偏推定量** 以下で定義される.

$$\widehat{\tau} = \sum_{g=1}^{G} q_g \widehat{\tau}_g \quad \text{where } q_g = \frac{n_g}{n}, \ \widehat{\tau}_g = \bar{Y}_{1g} - \bar{Y}_{0g}$$

*Proof.* 不偏性を示そう, 以下のように (全体の) ATE の不偏推定量と一致する.

$$\mathbb{E}[\widehat{\tau}] = \sum_{g=1}^{G} \frac{n_g}{n} \frac{1}{n_g} \sum_{i=1}^{n} G_{ig}[Y_i^*(1) - Y_i^*(0)] = \frac{1}{n} \sum_{i=1}^{n} \sum_{g=1}^{G} G_{ig}[Y_i^*(1) - Y_i^*(0)] = \frac{1}{n} \sum_{i=1}^{n} [Y_i^*(1) - Y_i^*(0)]$$

なお, 再三になるが期待値は処置割当に関してとられており, 潜在効果は確定的である. □

また分散は, 各層の独立性より $\mathbb{V}[\widehat{\tau}] = \sum_{g=1}^{G} q_g^2 \mathbb{V}[\widehat{\tau}_g]$ で与えられ, neyman 推定量を用いることで:

$$\widehat{\mathbb{V}}[\widehat{\tau}] = \sum_{g=1}^{G} q_g^2 \widehat{\mathbb{V}}_{\text{neyman},g}$$

が推定量として考えられる. **Ch 2.4** で見た neyman 推定量の性質より, 割当に起因する不確実性のみを考えるならば過剰推定で, 割当・抽出両方に起因する不確実性を考える場合には不偏推定量となる.

**■p.61 回帰分析と層化無作為化実験**　OLS 回帰によっても同様の推定量を得られる. しかし層割当ダミーに適当に条件付けし, パラメータが層によって制約されない状況を作り出す必要がある.

**定義: 飽和したモデル, Saturated Model**

層割当ダミーが処置割当と完全に交差し, 各層内の処置・統制群平均を完全に再現できるモデルのこと.

以下のようなモデルを指す: *3

$$Y_i = \sum_{g=1}^{G} \alpha_g G_{ig} + \underbrace{\sum_{g=1}^{G} \beta_g G_{ig} Z_i}_{\text{完全な交差項}} + \varepsilon_i.$$

各層 $g$ に対し以下が成り立つ, 条件付平均が完全に再現できていることが伺える.

$$\mathbb{E}[Y_i \mid G_{ig} = 1, Z_i = 0] = \alpha_g,$$
$$\mathbb{E}[Y_i \mid G_{ig} = 1, Z_i = 1] = \alpha_g + \beta_g,$$

したがって，OLS による層別処置効果の推定量は

$$\widehat{\beta}_g = \bar{Y}_{1g} - \bar{Y}_{0g},$$

すなわち層 $g$ における処置群と対照群の平均値の差に等しい. 一致性の証明は上の不偏性の証明で得た標本対応の形を用いれば確認できると思われる. これを加重平均してやれば集計推定量:

$$\widehat{\beta} = \sum_{g=1}^{G} q_g \widehat{\beta}_g$$

が ATE に不偏性と一致性を持つ推定量として求まる.

**■クラスター化無作為化実験**　クラスター化無作為化実験を行うときには, 先述した spillover effect の問題から, 興味のある推定対象として, **クラスターごとの ATE** と, **その単純平均** が考えられる.

**定義: クラスターごとの ATE とその単純平均効果**

$$\tau_g = \frac{1}{n_g} \sum_{i=1}^{n} G_{ig}[Y_i^*(1) - Y_i^*(0)], \quad \tau_C = \frac{1}{G} \sum_{g=1}^{G} \tau_g \tag{2.9}$$

後者はクラスターを個人と見なした完全無作為化実験として考えられ, 統計的分析もそれに従う. 前者に関しては以下の線形回帰モデルの OLSE によって推定が可能:

$$Y_i = \alpha + \beta Z_i + \varepsilon_i.$$

クラスター内標本では $Z_i$ が完全に相関しているため, **クラスター頑健標準誤差** を考える (**Ch 3.1.1**). *4

*3 定数項を落としているのでダミー変数の罠が回避されている.

*4 何をもって相関? $g$ と $Z_i$ の対応とか見ればそうだろうけど定式化が分からない.

**■クラスター化無作為化実験における相関構造**　クラスター化無作為化実験では, 個人ではなくクラスター $g = 1, \ldots, G$ を単位として処置が割り当てられる. クラスター $g$ に属する個体 $i = 1, \ldots, n_g$ に対しては, クラスター共通の処置指示変数 $Z_g \in \{0, 1\}$ が与えられ, 個人レベルの処置変数は $Z_i = Z_g$ で表される. このとき, 個人レベルの回帰モデル:

$$Y_i = \alpha + \beta Z_i + \varepsilon_i$$

を考えると, クラスター内では $Z_i = Z_g$ が同一であるため, $Z_i$ 同士が完全に相関しているとみなされるわけだが, この「$Z_i$ の相関」は, 次のように定式化できる. 個人レベルの処置ベクトル $\mathbf{Z}$ を以下のように定義する:

$$\mathbf{Z}^\top \coloneqq \left(Z_{11}, \ldots, Z_{n_1,1}, \ldots, Z_{1G}, \ldots, Z_{n_G,G}\right)$$

クラスター $g$ の処置変数 $Z_g$ は確率 $p = 1/{}_G\mathrm{C}_{G_1}$ で独立に割り当てられるため, クラスター内の共分散は:

$$\mathrm{Cov}(Z_{ig}, Z_{jg}) = \begin{cases} p(1-p) & (i, j \in g) \\ 0 & (i, j \notin g) \end{cases}$$

したがって, 全体の共分散行列は以下で与えられる:

$$\mathrm{Var}(\mathbf{Z}) = p(1-p) \underbrace{\begin{pmatrix} \mathbf{1}_{n_1}\mathbf{1}_{n_1}^\top & 0 & \cdots & 0 \\ 0 & \mathbf{1}_{n_2}\mathbf{1}_{n_2}^\top & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{1}_{n_G}\mathbf{1}_{n_G}^\top \end{pmatrix}}_{G\text{ のうち }G_1\text{ に }\mathbf{1}_{n_g}\mathbf{1}_{n_g}^\top\text{ が処置として入る}}$$

で与えられ, クラスター内では全ての $Z_i$ が完全に相関 (同一値) することが分かった.

**■p.66 政策判断基準**　「平均処置効果の推定値が正であれば, 執行のコストを無視すれば, その処置を政策として実現したほうがよい」という素朴な判断は, **ミニマックスリグレット基準** のもとで最適な意思決定. 共変量ごとに処置を選択するような判断を下したい際には, とくに共変量が連続なケースでは困難が生じる. これに対して **経験的厚生最大化** なるアプローチがあるらしい.

# 3　推定・検定の諸問題

**■p.69 クラスター頑健分散推定量**　末石先生の講義資料の方が分かりやすいと思います. 講義資料

*Proof.* 観測を $G$ 個のクラスターに分け，クラスター $g$ の観測数を $n_g$ とする．クラスター $g$ に対応する説明変数行列を $\mathbf{X}_g$（サイズ $n_g \times k$），残差ベクトルを $\mathbf{e}_g$（長さ $n_g$）とおくと:

$$\mathbf{X}^\top = \left(\mathbf{X}_1^\top, \mathbf{X}_2^\top, \ldots, \mathbf{X}_G^\top\right), \qquad \mathbf{e}^\top = \left(\mathbf{e}_1^\top, \mathbf{e}_2^\top, \ldots, \mathbf{e}_G^\top\right)$$

OLS 推定量の共分散行列（条件付期待値）は

$$V_{\hat{\boldsymbol{\beta}}}^{cluster} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\,\mathbb{E}[\mathbf{e}\mathbf{e}^\top \mid \mathbf{X}]\,\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}$$

で与えられる．中間の項 $\mathbf{X}^\top\mathbb{E}[\mathbf{e}\mathbf{e}^\top \mid \mathbf{X}]\,\mathbf{X}$ をクラスター分割で展開すると,

$$\mathbf{X}^\top\mathbb{E}[\mathbf{e}\mathbf{e}^\top \mid \mathbf{X}]\,\mathbf{X} = \left(\mathbf{X}_1^\top, \mathbf{X}_2^\top, \dots, \mathbf{X}_G^\top\right)\mathbb{E}\begin{bmatrix} \mathbf{e}_1\mathbf{e}_1^\top & \cdots & \mathbf{e}_1\mathbf{e}_G^\top \\ \vdots & \ddots & \vdots \\ \mathbf{e}_G\mathbf{e}_1^\top & \cdots & \mathbf{e}_G\mathbf{e}_G^\top \end{bmatrix}\begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_G \end{pmatrix}$$

$$= \sum_{g=1}^{G}\sum_{h=1}^{G}\mathbf{X}_g^\top\,\mathbb{E}[\mathbf{e}_g\mathbf{e}_h^\top \mid \mathbf{X}]\,\mathbf{X}_h.$$

異なるクラスター間の条件付け後の相関はゼロ, つまり

$$\mathbb{E}[\mathbf{e}_g\mathbf{e}_h^\top \mid \mathbf{X}] = \mathbf{O} \quad \text{for } g \neq h$$

を置くと，上の二重和は対角成分だけ残り,

$$\mathbf{X}^\top\mathbb{E}[\mathbf{e}\mathbf{e}^\top \mid \mathbf{X}]\,\mathbf{X} = \sum_{g=1}^{G} X_g'\,\mathbf{X}_g^\top\,\mathbb{E}[\mathbf{e}_g\mathbf{e}_g^\top \mid \mathbf{X}]\,\mathbf{X}_g.$$

さらに各クラスター内を展開しよう．クラスター $g$ の $i$ 番目の説明変数を $\mathbf{x}_{gi}$（長さ $k$）とすると,

$$\mathbf{X}_g^\top\,\mathbb{E}[\mathbf{e}_g\mathbf{e}_g^\top \mid \mathbf{X}]\,\mathbf{X}_g = \sum_{i=1}^{n_g}\sum_{j=1}^{n_g}\mathbb{E}[e_{gi}e_{gj} \mid \mathbf{X}]\,\mathbf{x}_{gi}\mathbf{x}_{gj}^\top.$$

以上をまとめて元の $V_{\hat{\beta}}^{cluster}$ に代入すると,

$$V_{\hat{\boldsymbol{\beta}}}^{cluster} = (\mathbf{X}^\top\mathbf{X})^{-1}\left(\sum_{g=1}^{G}\sum_{i=1}^{n_g}\sum_{j=1}^{n_g}\mathbb{E}[e_{gi}e_{gj} \mid \mathbf{X}]\,\mathbf{x}_{gi}\mathbf{x}_{gj}^\top\right)(\mathbf{X}^\top\mathbf{X})^{-1}. \tag{3.1}$$

これにより, 教科書の表現を得た. □

**■クラスターの規模**　でかいクラスターを考えるほど標準誤差は過剰推定される (検出力が下がる).

**■p.79 多重検定問題**　様々な種類の潜在結果を考える際に, ただ仮説検定を繰り返すだけでは, 本来制御の対象である検定のサイズを制御することが出来ないという問題.

**■p.83 ボンフェローニ検定**　多重検定問題の対策として, 和集合不等式に基づいて, それぞれの過誤率を $5/M\%$ とすることで族過誤率 (FWER) を 5% に抑える検定.

**■p.85 ボンフェローニ・ホルム検定**　明らかに帰無仮説から生成されていない検定統計量を逐次的に消去し棄却水準を下げて, 検出力をボンフェローニ検定よりも高める一方で, 族過誤率 (FWER) は依然 5% に抑える検定. 少なくとも一つの帰無仮説を棄却する確率自体はボンフェローニ検定と変わらない ($5/M\%$).

**■p.87 適切な制御対象と分析結果の利用目的**　ある処置が企業における複数の業績指標に及ぼす効果を評価するような場合, どの項目に差があるかによって異なる対応をとるような場合は族過誤率を気にする必要がある．だがある処置が企業の業績に何らかの影響があるかを見るだけならば気にする必要がなく, **偽検出比率**

**(FDP)** の期待値である **偽検出率 (FDR)** を考えれば十分である.

$$\mathrm{FDP} := \frac{\#\{i \in \{1,\ldots,n\}:\ H_i\ \text{が真かつ}\ H_i\ \text{が棄却される}\ \}}{\max\{\#\{i \in \{1,\ldots,n\}:\ H_i\ \text{を棄却する}\ \},1\}} \tag{3.5}$$

$$\mathrm{FDR} = \mathbb{E}[\mathrm{FDP}] \leq \alpha$$

■p.89 **ベンジャミン・ホッシュバーグ検定**　10% が水準として慣例となっている. 手順としては, 今までとは逆に, **棄却しにくい** ものから順に $p$ 値の検証を行い, 棄却された場合にはそれ以上に棄却しやすいものを全て棄却するような検定となっている. 例えば $j$ 番目に小さい $p_{(j)}$, すなわち $j$ 番目に棄却されやすい帰無仮説については, 棄却の基準を少し厳しくして, $p_{(j)} \leq j\alpha/M$ ならば $H_{(1)} \sim H_{(j)}$ まで全て棄却する. 検定統計量が独立である際には, 論理的には怪しいところがあるが, 相関構造に対してはある程度頑健なことを考慮して, よく用いられる検定といえる.

# 4　非遵守者

■p.92 **処置割当・受取**　割当 $Z_i$ に加え, 非遵守者を分析するため, $Z_i$ を受けた場合の **処置受取** を $D_i^*(Z_i)$ と表す. この章では処置受取も二値変数とする. 潜在結果モデル的に, 観測できる処置受取 $D_i$ は:

$$D_i = Z_i D_i^*(1) + (1 - Z_i) D_i^*(0)$$

潜在結果は $Y_i^*(Z_i, D_i^*(Z_i))$ のように改訂でき, 潜在結果の観測手段 (顕在結果) も以下のように修正される:

$$Y_i = D_i^*(Z_i) Y_i^*(Z_i, 1) + (1 - D_i^*(Z_i)) Y_i^*(Z_i, 0)$$

$Y_i^*(Z_i, D_i^*(Z_i))$ となるような潜在結果のみが観測できることを確認せよ.

■p.94 **潜在結果と処置**　潜在結果は処置受取には影響を受けるが, 処置割当には影響を受けないと考えられるのが自然. 教科書で示されている **処置受取の ATE** は:

$$\frac{1}{n}\sum_{i=1}^{n}[Y_i^*(\cdot,1) - Y_i^*(\cdot,0)] = \frac{1}{n}\sum_{i=1}^{n}[Y_i^*(D_i^*=1) - Y_i^*(D_i^*=0)]$$

となるため, 割当 $Z_i$ の如何に関わらず全て同じでありうる.

■p.95 **処置割当効果, ITT 効果**　処置を実際受け取るかに関わらず, 単に処置を割当することによる平均効果ならば, 非遵守者が存在する場合にも識別が可能である. これをベンチマークとして本章の議論は出発する. 順守率の割合は標本によって異なると考えられるため, この効果量は **外的妥当性に乏しく** 応用可能性が低い.

■p.95 **処置割当効果の不偏推定量**　**無作為化実験の仮定のもとで求められる**, と記載があるが, これを clarify しておこう. まず無作為化実験の定義について確認する:

**定義: 無作為化実験 (p.38)**

確率割当であり, 関数形が分析者にとって既知であるような処置割当メカニズム.

**定義: 古典的な無作為化実験 (p.38)**

無作為化実験のうち, 個別割当かつ条件付独立割当なもの. = 関数形が既知な正規処置割当メカニズム.

古典的な無作為化実験の例として, 完全無作為化実験や層化無作為化実験などが例として挙げられていた.

**p.45** では, **古典的な無作為化実験の下で, 平均値の差が ATE の不偏推定量になる** 事が示されており, *5 今回もこの議論を繰り返せば不偏推定量を得られることが分かる.

■p.96 **非遵守者がいる場合の処置割当メカニズム** 単に $\mathbf{Y}^*$ の種類が増えただけ. 注意すべきはそのあとで, **個別割当, 確率割当, 条件付独立割当** が置かれているようだ. すなわち, この章において無作為化実験の語を使う際には, **古典的な無作為化実験** (正規処置メカニズムである無作為化実験) として捉えるべきなようだ.

■p.97 **片側遵守者のケースでの識別** **表 4.1** を見よ. このケースでは, $D_i^*(1) = 1$ であるような個体を **遵守者**, $D_i^*(1) = 0$ であるような個体を **非遵守者** と定義できる. これは観測可能性を保証せず, 実際に $Z_i = 1$ であるような個人についてしかタイプを判別することが出来ない. 順守タイプの個体数 $n_G$ と 割合 $\pi_G$ は真の値であることに注意せよ.

■p.99 $\widetilde{\tau}_D$ **の解釈** 受取の差がある人のみ (遵守者) が足され $\pi_{com}$ と一致する, 変形は $D_i(0) \equiv 0$ による.

■p.99 **処置割当と潜在結果の条件付独立** **条件付独立割当** から得られる結果.

■p.99 **遵守タイプごとの結果の処置割当効果** $Z_i = 0$ となる個人は遵守タイプを観測することが出来ないので, これらについて不偏推定量を得ることはできない.

■p.100 **識別対象の明確化** 以下のように処置割当効果 (観測可能) は (観測不可能な) 推定量に分解できる.

$$\widetilde{\tau}_Y = \widetilde{\tau}_{Y,com}\widetilde{\tau}_D + \widetilde{\tau}_{Y,noc}(1 - \widetilde{\tau}_D)$$

我々が知りたいのは, 遵守者にとっての処置割当効果 (**観測不可能**) である $\widetilde{\tau}_{Y,com}$ である.

■p.100 **潜在的条件付独立** **条件付独立割当** から, 割当 $Z_i$ は潜在結果 $Y_i^*$ と条件付独立である. しかし $D_i$ とは条件付割当ではない. この設定下では以下の **潜在的な条件付独立** が成立することに注意する必要がある.

$$\mathbb{P}[D_i = 1 \mid Y_i^*, G_i = g] = \mathbb{P}[D_i = 1 \mid Y_i^{*\prime}, G_i = g] \quad \forall Y_i^*, Y_i^{*\prime} \tag{4.2}$$

$G_i$ に応じて証明が行われている. 更に, 片側非遵守における **遵守者にとっては** $D_i = Z_i$ が成り立つため , $D_i$ **に関する条件付独立も成立する**.

■p.101 **非遵守者に関する除外制約** $Y_i^*(Z_i, D_i^*(Z_i)) = Y_i^*(D_i^*(Z_i)) \quad \text{if } G_i = noc.$ $Z_i$ と潜在結果を関連させないためには, **二重盲検法** などによって処置割当が分からないようにするなどの操作が要求される. 非遵守者では $D_i^* \equiv 0$ より, $\widetilde{\tau}_{Y,noc} = 0$ で, 興味対象である $\widetilde{\tau}_{Y,com}$ は以下のように推定可能な形で表現できる:

$$\widetilde{\tau}_{Y,com} = \frac{\widetilde{\tau}_Y}{\widetilde{\tau}_D}, \qquad \widehat{\widetilde{\tau}}_{Y,com} = \frac{\widehat{\widetilde{\tau}}_Y}{\widehat{\widetilde{\tau}}_D} = \frac{\overline{Y}_1 - \overline{Y}_0}{\overline{D}_1 - \overline{D}_1}.$$

■p.102 **遵守者に関する除外制約** $Y_i^*(Z_i, D_i^*(Z_i)) = Y_i^*(D_i^*(Z_i)) \quad \text{if } G_i = com.$ この場合, **遵守者に対する平均処置効果, 局所平均処置効果 (LATE)** は $Z_i$ の値に関係なく, 遵守者の処置割当効果 $\widetilde{\tau}_{Y,com}$ と一致する:

$$\widetilde{\tau}_{Y,com} = \frac{1}{n_{com}} \sum_{G_i = com} (Y_i^*(\cdot, 1) - Y_i^*(\cdot, 0)) \tag{4.4}$$

*5 本文では完全無作為化実験を念頭に置いた記載だったが, 利用している仮定を見る限り古典的な無作為化実験では常に成立する結果だと考えられる.

本当に知りたかったのはこの LATE. 片側遵守者・各遵守タイプへの除外制約によって成立した結果であることは忘れてはならず, **実際にその仮定が成り立ちうるかは問題領域に基づいた慎重な検証が必要となる**.

■p.105 **両側非遵守者** $G \in \{nt, co, df, at\}$. **常時参加者・不参加者への除外制約** を仮定することで 2 タイプの処置割当効果は消える. さらに **反抗者の不存在** ($D_i^*(\cdot)$ の単調性) を仮定すれば片側遵守者と同様の議論に帰着できる. 最後の仮定に関しては **二重盲検法** などが要求される.

■p.108 LATE **の操作変数法としての解釈** **LATE が, 操作変数として $Z_i$ を用いた操作変数法における推定量として解釈でき, 2SLS 推定量によって推定・統計的推測を行えること** を示すのが目標である.

■**前提**: **内生性の問題点** **内生性が一致性を崩す** ことが問題である. 例えば以下の構造モデルを考える:

$$Y_i = \beta_0 + \beta_1 X_i + e_i \quad \text{with } Cov(X_i, e_i) \neq 0$$

ここで, OLSE $\widehat{\beta}_1$ の確率収束先は以下のようになる:

$$\widehat{\beta}_1 \xrightarrow{p} \beta_1 + \frac{Cov(X_i, e_i)}{Var(X_i)} \neq \beta_1$$

一致性が崩れており, これは $Cov(X_i, e_i) \neq 0$ なる内生性に起因した問題である. さらに以下の関係性:

$$\mathbb{E}[e_i \mid X_i] = 0 \implies \mathbb{E}[X_i e_i] = \mathbb{E}[e_i] = 0 \implies Cov(X_i, e_i) = 0$$
$$Cov(X_i, e_i) \neq 0 \implies \neg(\mathbb{E}[X_i e_i] = \mathbb{E}[e_i] = 0) \implies \mathbb{E}[e_i \mid X_i] \neq 0$$

から, この構造モデルが線形回帰モデルではなく, さらに線形射影モデルですらないことを明らかにする.

Ex. Professor Moore's law of demand 操作変数法を用いるモチベーション付けとして例を挙げよう. 需要・供給モデルでは, 需要量 $Q^d$ と供給量 $Q^s$ が価格 $P$ を介して同時に決定される. 同時方程式

$$\begin{aligned} \textbf{Demand} : Q^d &= \alpha + \underbrace{\beta}_{\text{theoretically} <0} P + \underbrace{\varepsilon_d}_{\text{Demand Shock}} \\ \textbf{Supply} : Q^s &= \gamma + \underbrace{\delta}_{\text{theoretically} >0} P + \underbrace{\varepsilon_s}_{\text{Supply Shock}} \end{aligned}$$

を考える. ここで $\varepsilon_d, \varepsilon_s$ はそれぞれ需要と供給の外生的ショック（誤差項）である. 市場均衡では $Q^d = Q^s$ となるが, 価格 $P$ は両方のショックに影響されるため, **価格と需要の誤差項 $\varepsilon_d$ は相関を持つ可能性が高い**. その結果, 需要方程式は OLS で一致推定不可能となってしまう.

需要・供給の同時決定下で観測された均衡点に対して単純に OLS 回帰を行うと, 需給が同時に変動するために価格と数量の間に正の相関が生じ, 誤って右上がりの需要曲線が推定されうる. 図 1 は観測された複数の均衡点（黒丸）から OLS で推定した場合に右上がりの需要曲線が得られる様子を示している.

図 2 は, 需要曲線と供給曲線がそれぞれ外生ショックによってシフトする場合の均衡点の変化を模式的に示している. 需要曲線 $D_0$（実線）が右方へシフトして $D_1$（破線）になると, 均衡点は $E_0$ から $E_2$ へと上昇する. 一方, 供給曲線 $S_0$ が右方へシフトして $S_1$ となると, 均衡点は $E_0$ から $E_1$ へと下落する. このように, 需要・供給ショックにより, 得られる均衡点 $(E_0, E_1, E_2)$ は右上がりの相関を示し得る. 図 2 の例では, 需要・供給ショックで異なる均衡点が生じ, それらを単純回帰すると先の図のような右上がりの需要曲線が得られる.

操作変数法では, 推定したい構造方程式（例では需要関数）を固定し, もう一方の供給曲線を外生変数で操作（シフト）しながら推定を行う. 具体的には, 需要曲線を固定し, 需要には無関係で供給にだけ影響する変数

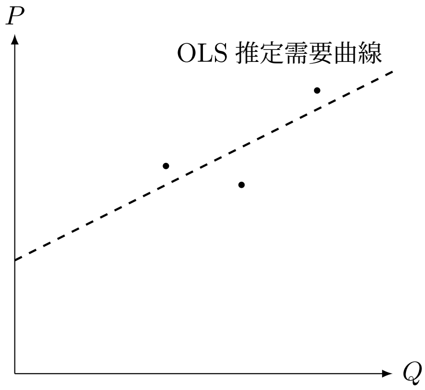


図 1　誤った推定による右上がりの需要曲線

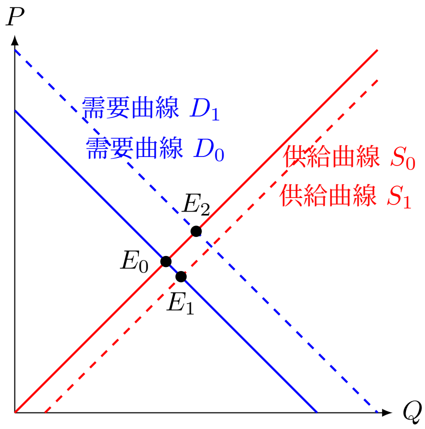


図 2　外生ショックによるシフト

$Z$（操作変数）を用いて供給曲線を右方へシフトさせる．こうして得られる複数の均衡点から需要曲線の傾きを推定する．手順を整理すると以下のようになる:

1. 推定対象の需要曲線 $D$ を固定し, 供給曲線 $S$ を外生的に右方へシフトさせる．
2. 操作変数 $Z$ は, 需要に影響せず供給に影響する変数（例: 自然災害や税金変更による供給ショック）．
3. 操作変数 $Z$ の変化により供給曲線が右方へシフトすると, 価格 $P$ と数量 $Q$ が変化し, 新たな均衡点 $(P, Q)$ が得られる．この変化を利用して需要曲線の真の傾きを推定できる．

価格 $P$ と取引量 $Q$ はモデル内で決定されるため**内生変数**である．一方, 需要ショック $\varepsilon_d$, 供給ショック $\varepsilon_s$, および操作変数 $Z$ は外部から与えられる**外生変数**である．**操作変数法では, 外生変数 $Z$ の変化を通じて内生的な価格変動を説明することで真の需要曲線を識別可能になる．**

**■末石 p.23 内生性と識別**　部分均衡分析の例でいえば, 需要ショックが価格に与える影響がある場合に, そのまま OLS で推定してもまともな結果が出ないことを思い起こせばわかる話だ．この識別の問題に対応するためには, 需要ショックを固定 $(Cov(Z, \varepsilon_d) = 0)$ する術が必要になる．

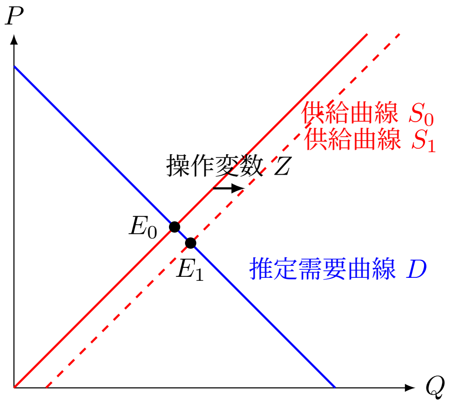


図 3　操作変数 $Z$ による供給曲線のシフト

**■末石 p.24 除外制約**　部分均衡分析の例なら, 操作変数 $Z$ は $Q^d$ に直接影響を与えない = 需要ショック $\varepsilon_d$ に含まれない必要が, 需要曲線を固定するためにある. まとめると, 以下の性質が操作変数には望ましい:

**操作変数に必要な性質**

1. **関連性**: $Corr(X_i, Z_i) \neq 0$
2. **外生性**: $Corr(\varepsilon_i, Z_i) = 0$

**■末石 p.24 二段階最小二乗法**　例に当てはめて考えることとしよう.

1. **第一段階**: 操作変数 $Z$ を用いて価格 $P$ を回帰し, $\widehat{P}_i = \widehat{\pi}_0 + \widehat{\pi}_1 Z_i$ を得る（誘導形）.
2. **第二段階**: 得られた予測値 $\widehat{P}_i$ を使って, 数量 $Q_i$ を需要方程式に回帰し, 需要弾力性 $\beta_1$ を推定する.

図 4 の通り, 誘導形での推定を挟み, 内生性の原因となる誤差項 $\nu_i = P_i - \widehat{P}_i$ を排除することによって, 需要ショックを変動させずに興味のある情報を取り出せる. 第一段階目は数学的に都合のいい部分だけ取り出しているもので, 実際の生成過程 (供給曲線) によらないことにも注意しよう.

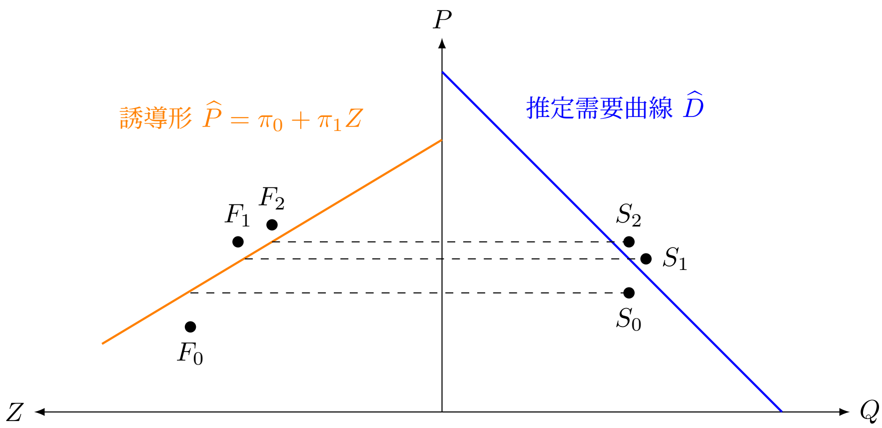


図 4　2SLS の流れ

**■p.108 観測式の問題点**　以下が成り立つ:

$$Y_i = \alpha + D_i \tau_{late} + \varepsilon_i + D_i \nu_i$$

$Y_i^*(Z_i, 0)$ のとき, 遵守者は $Y_i^*(0,0)$, 非遵守者は如何なる $Z_i$ でも成立するため, 辻褄合わせの為に $Z_i = 0$ で設定しこの結果が得られる. 問題点は, 相関 $(Cov(D_i, \varepsilon_i + D_i \nu_i) \neq 0)$ で一致性が崩れていること.

**■p.108 『無作為化実験の仮定』**　**p.54** でも利用されている, 『$Z_i$ が $Y_i^*$ と独立』が, ここでも意味されていると考えられる. これが成り立つ理由を考えよう.

**定義: 確率変数の独立性**

確率変数 $X, Y$ が $\mathbb{P}[X, Y] = \mathbb{P}[X]\mathbb{P}[Y]$ ならば, $X, Y$ は独立であるという.

**仮定: 条件付き独立割当**

$p_i$ が 潜在結果に依存せず, $\mathbb{P}[\mathbf{Z} \mid \mathbf{W}, \mathbf{Y}^*(0), \mathbf{Y}^*(1)] = \mathbb{P}[\mathbf{Z} \mid \mathbf{W}]$.

*Proof.* コントロール変数を落とせば, $\mathbb{P}[\mathbf{Z} \mid \mathbf{Y}^*(0), \mathbf{Y}^*(1)] = \mathbb{P}[\mathbf{Z}]$. つまり:

$$\mathbb{P}[Z_i, Y_i^*(\cdot)] = \mathbb{P}[Z_i \mid Y_i^*(0), Y_i^*(1)]\mathbb{P}[Y_i^*(\cdot)] = \mathbb{P}[Z_i]\mathbb{P}[Y_i^*(\cdot)]$$

で, 定義より $Z_i$ と $Y_i^*$ は独立である. □

これを利用することで成り立つのは:

$$\mathbb{E}[\varepsilon_i \mid Z_i] = 0$$

**■p.108 $\mathbb{E}[\varepsilon_i + D_i \nu_i \mid Z_i] = 0$ の成立**　無作為化実験の仮定は, 遵守タイプをコントロール変数とした条件付独立として利用され $\mathbb{E}[D_i \nu_i \mid Z_i] = 0$ が成り立つことと, $\mathbb{E}[\varepsilon_i \mid Z_i] = 0$ から以下が得られる:

$$\mathbb{E}[\varepsilon_i + D_i \nu_i \mid Z_i] = 0 \implies Cov(Z_i, \varepsilon_i + D_i \nu_i) = 0$$

即ち **除外制約 (外生性)** が満たされている. 片側遵守では明らかに **関連性** も満たされているため, **操作変数としての望ましい性質を $Z_i$ は満たしている**.

**■p.109 2SLS の手順**　更に変形が可能で, 2SLS の構造が明示化されている:

$$Y_i = \alpha + (\pi_0 + \pi_1 Z_i)\tau_{late} + \eta_i = \alpha + (\pi_0 + \pi_{com} Z_i)\tau_{late} + \eta_i$$

ここでは $\mathbb{E}[\eta_i \mid Z_i] = 0$ が成立している. 第一段階回帰: $D_i$ の $Z_i$ への回帰で,

$$D_i = \pi_0 + \pi_{com} Z_i + u_i, \quad \text{with } \mathbb{E}[u_i \mid Z_i] = 0.$$

$$\mathbb{E}[D_i \mid Z_i] = \pi_0 + \pi_{com} Z_i$$

第二段階回帰: $Y_i$ の $\widehat{D}_i$ への回帰で,

$$Y_i = \alpha + \tau_{late}\widehat{D}_i + \eta_i, \quad \text{with } \widehat{D}_i = \widehat{\pi}_0 + \widehat{\pi}_{com} Z_i.$$

$$\mathbb{E}[Y_i \mid Z_i] = \alpha + \tau_{late}(\pi_0 + \pi_{com} Z_i) = \alpha + \tau_{late}\pi_0 + \tau_{late}\pi_{com} Z_i$$

したがって推定量は

$$\widehat{\tau}_{late} = \frac{\widehat{\tau_{late}\pi_{com}}}{\widehat{\pi_{com}}} = \frac{\overline{Y}_1 - \overline{Y}_0}{\overline{D}_1 - \overline{D}_0} = \widehat{\tau}_{Y,com}. \tag{4.5}$$

繰り返しになるが, LATE は, (二値変数に限る) 操作変数として $Z_i$ を用いた操作変数法における推定量として解釈でき, 2SLS 推定量によって推定・統計的推測を行える.

# 5 無作為化実験の実践

■p.114 Baird et al. (2022) 要点は以下のようなものだろうか:

- 両側非遵守者のケースであることを確認.
- 他の職業訓練プログラムに応募した人は観測できていない.
- $c$ の書きぶりが $t$ にも依存していそう.
- 結果変数が二つあり, 多重検定問題は少し問題かもしれない.
- クラスターが過大であるゆえ保守的な結果が予期される.
- 両側非遵守者のケースにおける識別のための仮定が必要.
- 就労タイプを考えたければ標本を分割すべき.

■p.119 Aragon et al. (2020) 要点は以下のようなものだろうか:

- 層化無作為化実験にあたる.
- 与信に貸付方法が移行したことで長期効果は見れない.
- **示唆的証拠** による妥当性の強化
- 時間効果は $\mathbf{X}_{it}$ に吸収されている.
- グループ内での任意の共分散構造が許容されており $\#\{G_i\} = 3$.
- **表 5.4** は『統制群の結果の平均』のタイポと思われる.
- 途中で与信に切り替えた非遵守者は書きぶりからかなりの割合にのぼりそうだが, その妥当性は?
- 交差項の係数など, 直感に反する係数は説明すべき.
- 交差項の後者の影響が意味不明.
- 分析から得られた仮説も検証すべき, この際因果関係は特定できないので消極的な支持がせいぜい. しかし仮説として提示するには大事な姿勢.

■p.126 p-hacking 多重検定問題と一体.

■p.127 **事前登録制度** 事前査読制度と組み合わせて効力を発揮する.

■p.128 Afridi et al. (2021)

- スピルオーバー効果があるなら個人の ATE を見るのは不適切では?
- 無作為割当の前に抽出があるのが通常のクラスター化無作為化実験とは違う.
- 定式化の誤りをここで強調する理由は?
- 多重検定問題はより深刻になりうる.