Econometrics Note

Sora Maekawa *

2025 年 1 月 25 日

1 準備

1.1 確率論

確率空間 $(\Omega, \mathfrak{B}(\Omega), \mathbb{P})$ で数学的に定義される．ここで, $\Omega = \{\omega_1, \dots, \omega_M\}$, $M \in \mathbb{N}$ は標本空間（事象の集まり）,$\mathfrak{B}(\Omega) = \{A : A \subseteq \Omega\}$ は Ω のべき集合（確率を測りたい対象の集まり）で, これに対して $\mathbb{P} : \mathfrak{B} \to [0, 1]$ で確率を与える．

Ex. コイントス 結果を確率空間により数学的に記述すると以下のようになる．$M = 2$, $\Omega = \{\text{表}, \text{裏}\}$, $\mathfrak{B}(\Omega) = \{\phi, \{\text{表}\}, \{\text{裏}\}, \Omega\}$, $\mathbb{P}(\phi) = 0$, $\mathbb{P}(\{\text{表}\}) = \mathbb{P}(\{\text{裏}\}) = 1/2$, $\mathbb{P}(\Omega) = 1$.

確率変数 確率的に値を取る数のことで, 確率空間上の関数として定義される．ある確率変数 $X : \Omega \to \mathbb{R}$ に対し, 定義域, 値域それぞれの部分集合 $A \subseteq \Omega, B \subseteq \mathbb{R}$ を考える．この時, $X(A) = \{X(\omega) : \omega \in A\} \subseteq \mathbb{R}$ を X による A の像と呼び, $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} = \{X \in B\} \subseteq \Omega$ を X による A の逆像と呼ぶ. **NOTATION**: $\{X \in \{x\}\} \equiv \{X = x\}$. ここで x は X の実現値の一つ.

Ex. コイントスと賭け 前の例にてコインが表の時 100 ドルを, 裏の時 -50 ドルとなる賭けの損益を X とおく．このとき, X は確率変数であり, $X(\text{表}) = 100, X(\text{裏}) = -50$（$x \in \{100, -50\}$）．また, $\{X(\cdot) = x = 100\} = \{\omega \in \Omega : X(\omega) = 100\} = \{\text{表}\}$（裏でも同様に記せる）から $\mathbb{P}(X = -50) = \mathbb{P}(X = 100) = 1/2$ である．**NOTE**: 確率変数 X は事象 ω それ自体ではなく, ω の関数. 今回の場合ならコインの面ではなく, それによる損益.

条件付き期待値の諸性質 $\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X))$ は **LIE（繰り返し期待値の法則）**と呼ばれ, 更に $\mathbb{E}(Y|X) = \mathbb{E}(\mathbb{E}(Y|X, Z)|X)$（中にある期待値ほど情報集合が小さいことも確認しよう）．他の変形も確認しよう, 一般に, X と W を $X = f(W)$ の関係性をもつ確率変数とすると,

* 一橋大学経済学部 3 年, 五年一貫専修コース公共経済プログラム, 佐藤ゼミ 24 期

$\mathbb{E}(Y|X) = \mathbb{E}(\mathbb{E}(Y|W)|X)$ である．$X = f(W) = W$ のときには, $\mathbb{E}(Y|X) = \mathbb{E}(\mathbb{E}(Y|X)|X)$ であることにも注意しよう（回帰の性質を導く際に用いる）．また, $\mathbb{E}(g(X)Y|X) = g(X)\mathbb{E}(Y|X)$ や, 全分散の法則 $Var(Y|X) = \mathbb{E}(Var(Y|X)) + Var(\mathbb{E}(Y|X))$ も成立する．

Ex. 共分散　二つの確率変数 X, Yが $P(Y = y|X = x) = P(Y = y)$, つまり独立であるときのことを考える．この場合 $\mathbb{E}(Y|X) = \mathbb{E}(Y) = \mu_Y$. つまり, Yの条件付き期待値が X に依存しないので, 共分散 $Cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ は, $\mathbb{E}(XY) = \mathbb{E}(\mathbb{E}(XY|X)) = \mathbb{E}(X\mathbb{E}(Y|X)) = \mathbb{E}(\mu_Y X) = \mu_Y\mathbb{E}(X) = \mathbb{E}(X)\mathbb{E}(Y)$ より $Cov(X, Y) = 0$ となる. **NOTE**:LIE の派生形を利用した．

確率論の諸定理　イェンセンの不等式 (Jensen's Inequality): $g(X)$ が凸関数ならば, $g(\mathbb{E}[X]) \leq (\mathbb{E}[g(X)])$. 特に $g(X) = |X|$ に関しこの公式は Expectation Inequality と呼ばれる．コーシーシュワルツの不等式 (Cauchy-Schwarz Inequality): $|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$ チェビシェフの不等式 (Chebychev's Inequality): Yを確率変数, c を定数とすると, $P(|Y - c| \geq \epsilon) \leq \frac{\mathbb{E}[(Y-c)^2]}{\epsilon^2}$ $\forall \epsilon > 0$. チェビシェフの不等式は LLN の証明に利用できることもよく知られている．

1.2 統計的推測

用語の定義　n 個の確率変数 $\{\mathbf{X}_i\}_{i=1}^n : \Omega \to \mathbb{R}^d, \overset{i.i.d.}{\sim} F$ を考える．このとき, F を**母集団分布**, $\{\mathbf{X}_i\}_{i=1}^n$ をサイズ n の**無作為標本**とよぶ．Fの平均, 分散は**母平均, 母分散**と今後表記する．実際に我々が得られるデータ $\{\mathbf{x}_i\}_{i=1}^n = \{\mathbf{X}_i(\omega)\}_{i=1}^n$ を**実現値**, または**観測値**と呼ぶ. $\{\mathbf{X}_i\}_{i=1}^n$ で構成される Ω 上のランダムな関数を**推定量**と呼び, その**実現値**を **推定値**という．推定量 $\hat{\Theta}_n = \hat{\Theta}_n(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ が推定したい真のパラメータ $\Theta_n = \Theta_n(F)$ を**被推定量**という．**統計量**とは, 標本データから計算される任意の量 (Ex. 標本最大値, 中央値) を指し, 必ずしも母集団パラメータを推定するもの（＝推定量）に**限らない**, ガウス統計量, χ^2 統計量, t 統計量, F統計量などが知られている．その中でも**検定統計量**とは, 仮説検定において, **帰無仮説のもとで**分布が特定され, 棄却域の構成に用いられる確率変数のことを指す．仮説検定で犯しうる誤りは, 帰無仮説を誤って棄却する (タイプⅠエラー) ものと帰無仮説を誤って採択する (タイプⅡエラー) ものの二つである．仮説検定に先立ち設定される, **許容できるタイプⅠエラーの確率** $\alpha = P(\text{rejecting } H_0|H_0 \text{ is ture})$ こそが**有意水準**である. 慣習的に選ばれることが多いのは $\alpha = 0.05$ である. 母集団から同じ方法で 100 回ランダムに標本を抽出して信頼区間を構成した場合, 約 $100(1 - \alpha)$ 回 (× 確率○割合) 信頼区間が真の平均を含むと期待される. 母平均は非確率的であり, 含むか含まないのどちらかのみであるため, 信頼区間については頻度論から議論を行う必要があることに注意せよ．最後に, p 値とは『**帰無仮説の下で, 実際の実現値と同等かそれ以上に, その帰無仮説に不利な実現値を検定統計量が出す確率**』を指す.

Ex. 標本平均　$F : \mathbb{R} \to [0, 1], X, X_1, \ldots, X_n \overset{i.i.d.}{\sim} F$ とする．$\mu = \mathbb{E}[X]$, $\sigma^2 = Var(X)$ とし, これを推定するため（被推定量としたとき）の**推定量**（確率変数、推定値ではない）を $\hat{\mu} = \hat{\mu}(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^n X_i$ として, これを標本平均 $\bar{X_n}$ と定義する．この実現値

$\hat{\mu} = \hat{\mu}(x_1, \dots, x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$ が推定値で, これは非確率変数. また, $\mathbb{E}[\hat{\mu}] = \mu$ で不偏推定量である. なお, 標本平均の分散は $Var(\bar{X}_n) = \frac{\sigma^2}{n}$ であり, この平方根は標準誤差 (Standard Error, SE) という. 標本平均のみならず, 他推定量の分散の平方根についても標準誤差と呼ぶ.

Ex. p 値の計算　ソムリエはワインの赤白をその鋭い味覚で感じ取れると言われている. 対して我々はソムリエは違いなど一切分かっていないと疑いの目を向ける. そこで, 彼/彼女に目隠しをつけ, 赤と白の二択でワインを 10 回区別してもらった. 実際に 10 回のうち 9 回は二択を当てることが出来たときの, 帰無仮説『彼/彼女はあてずっぽうで色を言っているだけ』の下での p 値は？
解答: ” 帰無仮説の下での分布” は二項分布である. 実現値は 9 回成功ゆえ, これがあてずっぽうで実現する確率は $10C1 \times (1/2)^{10}$. では, ” 実際と同等かそれ以上に帰無仮説に不利な実現値をだす確率” とは何だろうか? まず, 帰無仮説に不利な実現値はソムリエの成功回数が多い結果のことを指す. つまり今回求めたい確率 (= p 値) は 9 回以上 (＝実際の実現値以上に) 成功する確率. つまり:

$$p\text{値} = 10C0 \times (1/2)^{10} + 10C1 \times (1/2)^{10} = 11/1024$$

Ex. 標本平均の信頼区間　帰無仮説の仮定下では, $P(-c \leq t_\mu \leq c) = 1 - \alpha$ であり, 左辺を変形すると, $P(-c \leq t_\mu \leq c) = P(-c \leq \sqrt{n}\frac{\bar{X}_n - \mu_{X,0}}{\sigma_X} \leq c) = P(\bar{X}_n - c\frac{\sigma_X}{\sqrt{n}} \leq \mu_{X,0} \leq \bar{X}_n + c\frac{\sigma_X}{\sqrt{n}})$ であるから, $[\bar{X}_n - c\frac{\sigma_X}{\sqrt{n}}, \bar{X}_n + c\frac{\sigma_X}{\sqrt{n}}]$ が $100(1-\alpha)$ パーセント信頼区間である. ここで臨界値である c や分散推定量である $\frac{\sigma_X}{\sqrt{n}}$ は推定の (漸近) 分布と状況によって変わることに注意せよ. 標本が十分に大きいときの標本平均について 95 パーセント信頼区間を構成するならば, $c \approx 1.96$ である.

Ex. t 統計量を用いた両側検定　$X_1, \dots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$ である. 先ず, σ_X^2 が既知として, $H_0 = \mu_{X,0}, \alpha = 0.05$ の両側検定を考える. ここで, **実際に帰無仮説が正しければ**, t 検定統計量は $T = \sqrt{n}\frac{\bar{X}_n - \mu_{X,0}}{\sigma_X} \sim \mathcal{N}(0,1)$ で標準正規分布に従う. 次に, σ_X^2 も未知として同様の両側検定を考える. ここで, t 統計量は $T = \sqrt{n}\frac{\bar{X}_n - \mu_{X,0}}{s_X} = \frac{\sqrt{n}\frac{\bar{X}_n - \mu_{X,0}}{\sigma_X}}{\sqrt{\frac{(n-1)s_X^2/\sigma_X^2}{n-1}}} \sim t(n-1)$ で, 自由度 $n-1$ の t 分布に従う.

漸近理論　確率変数 X の母集団分布が分かっていないこともしばしばである. その場合, 我々が標本平均などの推定量を構成する意味とは? ここで重要になるのが標本数 n を無限大に大きくした際の統計量の近似的な性質を探る漸近理論, とくに大標本理論である. $\lim_{n\to\infty} P(|z_n - c| \geq \epsilon) = 0 \quad \forall \epsilon > 0$ なら, 数列 $\{z_n\}$ は c に**確率収束**するという. $z_n \overset{p}{\to} c$ かつ $y_n \overset{p}{\to} d$ ならば, $z_n + y_n \overset{p}{\to} c + d$ かつ $z_n y_n \overset{p}{\to} cd$. $\{S_n\}$ は確率変数列で, F_n はこの cdf として, F_n が全ての F(S の cdf) 上の連続な点で F へ収束するなら, $\{S_n\}$ は S に**分布収束**するという, この時の F が S_n の**漸近分布**である. 大数の法則 (LLN) は確率収束を, 中心極限定理 (CLT) は分布収束を用いて定義され, どちらも漸近理論の枠組みでの結果である.

漸近理論の諸定理　スラツキーの定理 (Slutsky's Theorem): $z_n \overset{p}{\to} c$ かつ $S_n \overset{d}{\to} S$ ならば, $z_n + S_n \overset{d}{\to} c + S$, $z_n S_n \overset{d}{\to} cS$, かつ $S_n / z_n \overset{d}{\to} S/c$ if $c \neq 0$. 連続写像定理 (Continuous Mapping

Theorem): g が連続関数のとき, $z_n \xrightarrow{p} c$ ならば, $g(z_n) \xrightarrow{p} g(c)$, $S_n \xrightarrow{d} S$ ならば, $g(S_n) \xrightarrow{d} g(S)$. 大数の法則 (Law of Large Numbers, LLN): $X_1, \dots, X_n \overset{i.i.d}{\to} F(\mu_X, \sigma^2 < \infty)$ ならば, $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{p} \mu_X$. 中心極限定理 (Central Limit Theorem, CLT): $X_1, \dots, X_n \overset{i.i.d}{\to} F(\mu_X, \sigma_X^2 < \infty)$ ならば, $\sqrt{n}\frac{\bar{X_n}-\mu_X}{\sigma_X} \xrightarrow{d} \mathcal{N}(0,1)$ である. CLT については, 標本平均の標準化が行われており, それ故に標本平均の分散 $\frac{\sigma^2}{n}$ が用いられてこの形になっていることに注意せよ.

Ex. 標本分散・共分散 $F : \mathbb{R} \to [0,1]$ $X, X_1, \dots, X_n \overset{i.i.d.}{\sim} F$. 母集団平均 $\mu = \mathbb{E}[X]$, 母集団分散 $Var(X) = \sigma^2$ とし, **推定量**を $s^2 = s^2(X_1, \dots, X_n) = \frac{1}{n-1}\sum_{i=1}^n (X_i - \hat{\mu})^2$ として, これを標本分散 s^2 と定義. この実現値 $s^2 = s^2(x_1, \dots, x_n) = \frac{1}{n-1}\sum_{i=1}^n (x_i - \hat{\mu})^2$ が推定値である. $\mathbb{E}[s^2] = \sigma^2$ で不偏推定量である. また, LLN と連続写像定理より $s^2 \xrightarrow{p} \sigma^2$ で一致推定量であることも分かる (証明は他資料参照). 標本平均の標準誤差の推定量は, $SE(\bar{X}_n) = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$ と標本分散を利用している. 標本共分散も同様に, (二つの確率変数 (X, Y) について)$\frac{1}{n-1}\sum_{i=1}^n (X_i - \hat{\mu_X})(Y_i - \hat{\mu_Y})$ と構成すれば $Cov(X, Y)$ の不偏かつ一致推定量となる事が知られている.

Ex. OLSE と標本対応 通常の単回帰モデルに従い生成された n 組の i.i.d 標本 $(X_i, Y_i)_{i=1}^n$ を用いて OLS 推定を行うと, 傾きの推定量 $\hat{\beta_1} = \frac{Cov(X,Y)}{Var(X)}$ を得られる (後述). この**標本対応**は $b_1 = \frac{s_{X,Y}}{s_X^2}$ だが, この β_1 から b_1 への変換に用いた標本対応に正統性はあるのだろうか? X の母集団分散 $Var(X) = \sigma^2$, 母集団共分散 $Cov(X, Y) = \sigma_{X,Y}$ に対する不偏・一致推定量として標本分散 s_X^2, 標本共分散 $s_{X,Y}$ がある. ここで, 確率収束の性質と連続写像定理より, $\frac{s_{X,Y}}{s_X^2} \xrightarrow{p} \frac{\sigma_{X,Y}}{\sigma^2}$. 不偏性は期待値を取れば確認できる. 以上より, 標本対応の正統性を不偏性と一致性の観点から確認できた.

Ex. t 統計量を用いた両側検定 (Cont.) $X_1, \dots, X_n \overset{i.i.d.}{\sim}$, $\mathbb{E}[X_i^4] < \infty$ だが分布が不明な場合の t 統計量を考える. $T = \sqrt{n}\frac{\bar{X_n}-\mu_{X,0}}{s_X}$ を先述のように分解すると, (分母) は 1 に確率収束し, (分子) は $\mathcal{N}(0,1)$ に分布収束する ($\because$ CLT). Slutsky's Theorem より, $T = \sqrt{n}\frac{\bar{X_n}-\mu_{X,0}}{s_X} \xrightarrow{d} \mathcal{N}(0,1)$, つまり **t 統計量の漸近分布は標準正規分布**. 標本サイズが大きい限り, t 統計量の分布は標準正規分布で良く近似できることがわかった.

1.3 行列

NOTATION: 行列 $\mathbf{A}$ の転置行列を $\mathbf{A}'$ と, k 次単位行列を $\mathbf{I}_k$ と書く. n 次対称行列 $\mathbf{A}$, n 次ベクトル $\mathbf{x}$ を考えたとき, $\mathbf{x}'\mathbf{A}\mathbf{x}$ を $\mathbf{x}$ の **2 次形式（quadratic form）**とよぶ. 全ての $\mathbf{x} \neq \mathbf{0}$ について $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ ($\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$) なら, $\mathbf{A}$ は **正値定符号行列（非負値定）(positive (semi)definite)**と呼び, $\mathbf{x}'\mathbf{A}\mathbf{x} < 0$ ($\mathbf{x}'\mathbf{A}\mathbf{x} \leq 0$) なら, **負値定符号行列（非正値定）(negative (semi)definite)**と呼ぶ.

階数に関する諸性質 正則行列 $\mathbf{A}$ には, $\mathbf{A}\mathbf{c} = \mathbf{0}$ の解が $\mathbf{c} = \mathbf{0}$ しか存在せず, また逆行列 $\mathbf{A}^{-1}$ が存在する. どの $\mathbf{A}$ にも, $rank(\mathbf{A}) = rank(\mathbf{A}\mathbf{A}') = rank(\mathbf{A}'\mathbf{A})$ が成立する. $\mathbf{x}'\mathbf{A}\mathbf{x} \lneqq 0$ ならば $\mathbf{A}$ は正則.

多変量解析と行列　n 次元確率ベクトル $\mathbf{X}' = (X_1, \dots, X_n)$ を考える．この確率ベクトルの期待値は期待値のベクトルであり, 以下のように表せる．

$$\mathbf{E}[\mathbf{X}] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_n] \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = \mathbf{M}$$

共分散行列は $\Sigma = \mathbf{Var}(\mathbf{X}) = \mathbf{E}((\mathbf{X}-\mathbf{M})(\mathbf{X}-\mathbf{M})')$ で与えられる．$Cov(X_i, X_j) = Cov(X_j, X_i)$ から, 共分散行列は対称行列であり, また, $\Sigma = \mathbf{E}(\mathbf{X}\mathbf{X}') - \mathbf{M}\mathbf{M}'$ である．次に, 平均ベクトル $\mathbf{M}$, 共分散行列 Σ, そして定数ベクトル $\mathbf{a}' = (a_1, \dots, a_n)$ を持つ n 次元確率ベクトル $\mathbf{X}' = (X_1, \dots, X_n)$ を考える．$z = \mathbf{a}'\mathbf{X} = a_1X_1 + \cdots + a_nX_n$ は確率変数（スカラー）で, z の平均は $\mathbb{E}(z) = a_1\mathbb{E}(X_1) + \cdots + a_n\mathbb{E}(X_n) = \mathbf{a}'\mathbf{E}(\mathbf{X}) = \mathbf{a}'\mathbf{M}$ で与えられる．分散は, $Var(z) = Var(\mathbf{a}'\mathbf{X}) = \mathbb{E}((\mathbf{a}'\mathbf{X} - \mathbf{a}'\mathbf{M})(\mathbf{a}'\mathbf{X} - \mathbf{a}'\mathbf{M})') = \mathbb{E}(\mathbf{a}'(\mathbf{X}-\mathbf{M})(\mathbf{X}-\mathbf{M})'\mathbf{a}) = \mathbf{a}'\mathbf{E}((\mathbf{X}-\mathbf{M})(\mathbf{X}-\mathbf{M})')\mathbf{a} = \mathbf{a}'\mathbf{Var}(\mathbf{X})\mathbf{a} = \mathbf{a}'\Sigma\mathbf{a}$（$\Sigma$ の 2 次形式）となる．$Var(z) \geq 0$ ゆえ, この 2 次形式は非負値定符号行列. 更に, 線形結合の集合を考えてみよう．平均ベクトル $\mathbf{M}$, 共分散行列 Σ, そして $n \times K$ 定数行列 $\mathbf{A}$ を持つ n 次元確率ベクトル $\mathbf{X}' = (X_1, \dots, X_n)$ を考える．$\mathbf{z} = \mathbf{A}'\mathbf{X}$ は確率ベクトル（$\times$ スカラー）で, 以下のようにも表記できる．

$$\mathbf{z} = \mathbf{A}'\mathbf{X} = \begin{pmatrix} a_{1,1}X_1 + \cdots + a_{n,1}X_n \\ \vdots \\ a_{1,K}X_1 + \cdots + a_{n,K}X_n \end{pmatrix}$$

先ほどと同様に, z の平均は $\mathbf{E}(\mathbf{z}) = \mathbf{A}'\mathbf{M}$ で, 分散は $\mathbf{Var}(\mathbf{z}) = \mathbf{A}'\Sigma\mathbf{A}$ で与えられる．ベクトル, 行列での微分も可能である．$\frac{\partial g(\mathbf{x})}{\partial \mathbf{x}}$ は, $n \times 1$ ベクトルで各 x_i 成分での偏微分導関数が対応する．特に, n 次ベクトル $\mathbf{a}$ と n 次正方行列 A について, $\frac{\partial}{\partial \mathbf{x}}\mathbf{a}'\mathbf{x} = \mathbf{a}$, $\frac{\partial}{\partial \mathbf{x}'}\mathbf{A}\mathbf{x} = \mathbf{A}'$, $\frac{\partial}{\partial \mathbf{x}}\mathbf{x}'\mathbf{A}\mathbf{x} = (\mathbf{A} + \mathbf{A}')\mathbf{x}$. 最後に行列の表記を用いて多変数での CLT を挙げる．Multivariate Central Limit Theorem (CLT, Lindeberg-Levy):　$\{\mathbf{z}_i\}$ は i.i.d, 平均 $\mathbf{E}[\mathbf{z}_i] = \mu$ で, 共分散行列 $\mathbf{Var}(\mathbf{z}_i) = \Sigma$ は正値定符号行列かつ有限とする．このとき, $\sqrt{n}(\bar{\mathbf{z}}_n - \mu) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma)$.

2　回帰分析の基礎

2.1　回帰分析の必要性

前提として, 我々の関心は, ある施策によりもたらされた帰結＝**因果効果の推定**と, 現状のデータを用いた**将来の予測**の 2 つにある（予測のみが目標ならば, 因果効果を特定する必要はない）．理想的な因果推論のための実験として, **Randamized Controlled Trial（RCT）**がある．

ベンチマーク:RCT　我々が関心を持つ対象が, X が一単位変化したときの Y への**因果効果**だとする．RCT では, $X = 1$ の処置群と $X = 0$ の処置群を**ランダム**に振り分け, グループ間での差異はその処置によるものだけである状況を作り出す．この場合, 処置による**因果効果**は $\mathbb{E}[Y|X = 1] - \mathbb{E}[Y|X = 0]$ で与えられる．しかし, 我々が入手可能なデータには, 実験データと観測データがあり, 特に観測データにおける処置はランダムに割り当てられているとは限らず, 処置の因果効果を関連要素から切り離して見出すことは困難になる（後述）．ここからはランダムな割り当てが保証されない場合の予測と因果推論についての考察を進める．

因果推論と条件付き期待値　適切な (疑似相関の問題等を考慮した) 変数選択の下でコントロールを行い (→コントロール変数), **ceteris paribus・ほかの条件全て一定**の状況を（仮想的に）作り出し, **比較静学**への導入を行うのが, 特に因果推論の文脈における計量経済学のツールの役割といえる．平均, または期待される反応に焦点を当てるなら、比較静学は、$\mathbb{E}(Y|W, \mathbf{C})$, つまり条件付き期待値の推定を必要とする．ここで, ベクトル $\mathbf{C}$ は, Y の期待値に対する W の効果を調査するとき, 明示的に**固定したいコントロール変数の集合**を示す．$\mathbf{C}$ が上手くコントロールされていれば, W が連続のとき, **部分効果** $\frac{\partial \mathbb{E}(Y|W,\mathbf{C})}{\partial W}$ が直接的に因果効果といえよう．しかし $\mathbf{C}$ は観測不可能なことや, 効果を測るための適切な尺度がないこともしばしばである．このような理由から欠落変数バイアスが生じ, 内生性が存在したり (→ IV 法), そもそも因果関係を逆に取り違えたり, 因果関係が実際には見せかけだった, という問題が因果推論にはついて回ることに注意せよ．

Ex. 教育の賃金への因果効果　賃金への影響を与える要因が, 教育年数, 就業年数, そして能力のみであると考えよう．このとき求めるべき部分効果は, $\frac{\partial \mathbb{E}(wage|edu,exper,abil)}{\partial edu}$ で, コントロール変数ベクトル $\mathbf{C} = (exper, abil)$ だが能力は**観測不可能**.

予測と条件付き期待値　一方で, 多くの応用計量経済学での目標は, 説明変数ベクトル $\mathbf{X}$ で条件付けた内生変数 Y の期待値 $\mathbb{E}(Y|\mathbf{X})$ の**推定, 予測, 仮説検定（× 因果推論）**にある．Y の期待値が発散しなければ, $\mathbb{E}(Y|\mathbf{X}) = \mu(\mathbf{X})$ を満たすような関数 $\mu = \mu(\mathbf{X})$ が存在して, これが Y の $\mathbf{X}$ による平均変化をしめす．経済学では, 条件付き期待値は通常, 有限個のパラメータに依存するように設定される (=**パラメトリックモデル**), μ の自由度が下がることに注目.

Ex. $K = 2$ 説明変数モデル

$$\mathbb{E}(Y|\mathbf{X}) = \beta_1 + \beta_2 X_2 + \beta_3 X_3 \tag{1}$$

$$\mathbb{E}(Y|\mathbf{X}) = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_3^2 \tag{2}$$

$$\mathbb{E}(Y|\mathbf{X}) = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_2 X_3 \tag{3}$$

$$\mathbb{E}(Y|\mathbf{X}) = \exp[\beta_1 + \beta_2 \log X_2 + \beta_3 X_3], y \geq 0, x_2 > 0 \tag{4}$$

上三つは**パラメータに対して線形**, 最後の一つは非線形（$\because \exists \exp$）. 違いは部分効果と弾力性を考慮することで明らかになる（後述）.

μ を微分可能な変数としてみなせば, $\Delta\mathbb{E}(Y|\mathbf{X}) \approx \frac{\partial \mu(\mathbf{X})}{\partial X_i} \cdot \Delta X_i$ の関係が成立する. 実際に X が連続の時の部分効果は前回と同様 $\frac{\partial \mathbb{E}(Y|\mathbf{X})}{\partial X_i}$ で, 離散の場合には差分で推定される. 弾力性は $\frac{\partial \log \mathbb{E}(Y|\mathbf{X})}{\partial \log X_i}$.

Ex. $K = 2$ **説明変数モデル（Cont.）** 連続として考える. それぞれで部分効果を求めると, (1) 式では, (β_2, β_3), (2) 式では, $(\beta_2, \beta_2 + 2\beta_3 x_3)$, (3) 式では, $(\beta_2, \beta_2 + 2\beta_3 x_3)$, (4) 式は複雑に. 関数形で部分効果に他説明変数が介在する状況を表現できると分かる. 一方で, (4) 式の X_2 に対する弾力性は β_2 で一定. 弾力性推定における両辺 log の変換の有効性, $\exp^{-1}(\cdot) = \log(\cdot)$ による結果.

2.2 回帰モデルとは

回帰とは, **被説明変数を説明変数ベクトルによって予測**することである. 条件付き期待値 $\mathbb{E}[Y|X_2]$ は Y**の** X_2 **への回帰**と呼ばれ, 回帰モデルは以下のように表現できる.

$$Y = \mathbb{E}[Y|X_2] + (Y - \mathbb{E}[Y|X_2]) = \mathbb{E}[Y|X_2] + \epsilon$$

ここで導入した**誤差項（error term）** $\epsilon = Y - \mathbb{E}[Y|X_2]$ は, 両辺に X_2 での条件付き期待値をつけることで, $\mathbb{E}[\epsilon|X_2] = \mathbb{E}[Y|X_2] - \mathbb{E}[\mathbb{E}[Y|X_2]|X_2] = 0$($\because$ LIE の変形). ゆえに, $\mathbb{E}[\epsilon|X_2] = 0$ の性質を満たすことが分かる. 以後 LIE とその変形より, $\mathbb{E}[\epsilon] = 0$, $\mathbb{E}[X_2\epsilon] = 0$, $\mathbb{E}[h(X_2)\epsilon] = 0$ を得る.

最小二乗法 条件付き期待値の, Y の予測としての望ましさを確認したい. ここで用いるのは**最小二乗法（OLS）**である. 確率変数 X の関数である, Yの推定量を $g(X)$ とおく. このとき**推定誤差**は $e = Y - g(X)$ で与えられ, 以下の平均二乗誤差（MSE）を最小化するような推定量を考える.

$$Loss = \mathbb{E}[(Y - g(X_2))^2]$$

これを最小化するのは, $g(X_2) = \mathbb{E}[Y|X_2]$ であり, 最適な推定量が条件付き期待値と合致することが分かる. 尚, このときの予測誤差 e は, $e = Y - g(X) = Y - \mathbb{E}[Y|X_2]$ であることから, 誤差項 ϵ と一致することも分かる. ノンパラメトリック回帰は, それぞれの条件付期待値が被説明変数の推定量となる. しかし, この回帰について我々は, データ全体での傾向読み取りが困難であるなどの課題点を学んだ (R による実証分析 4 章参照).

2.3 単回帰モデルの導入

先述の回帰モデルにおける条件付き期待値の特定化を図る. 線形単回帰モデルでは, 実際に条件付き期待値が $\mathbb{E}[Y|X_2] = \beta_1 + \beta_2 X_2$（＝標本が線形関係に従い生成される）と仮定する.

β の推定　ここで, 無作為標本 $(X_{1,2}, Y_1), \dots, (X_{n,2}, Y_n)$ が得られたとして, 以下を考える.

$$Y_i = \beta_1 + \beta_2 X_{i,2} + \epsilon_i \quad i = 1, \dots, n$$

ここで, Y_i は従属変数, 内生変数, $X_{i,2}$ は独立変数, 外生変数, 説明変数などとよばれ, ϵ_i はランダムで観測不可能な**誤差項（error term）**である. この場合にデータの予測を行いたければ, (β_1, β_2)（未知）を推定すればよい. では考えうる推定方法は?　**NOTE**: 単回帰モデルの仮定下では, 誤差項は説明変数の変化を経由しない従属変数変動の要因を全て内包している.

単回帰モデルにおける最小二乗法　一般の回帰モデルと同様に, 先述した最小二乗法を単回帰モデルに適応すれば良い. 上の議論より, $g(X_2) = \mathbb{E}[Y|X_2]$ である. $\mathbb{E}[Y|X_2] = \beta_1 + \beta_2 X_2$, $Y_i = \beta_1 + \beta_2 X_{i,2} + \epsilon_i \quad i = 1, \dots, n$ の場合, 関数形を決定するパラメータとして選択できるのは (β_1, β_2) の 2 つ. (β_1, β_2) の OLS 推定量 $(\hat{\beta}_1, \hat{\beta}_2)$ は, 以下の MSE の minimizer である.

$$Loss = \mathbb{E}[e^2] = \mathbb{E}[(Y - g(X_2))^2] = \mathbb{E}[(Y - \beta_1 - \beta_2 X_2)^2]$$

FOC で得た正規方程式を用い, OLS 推定量 $(\hat{\beta}_1, \hat{\beta}_2) = (\mathbb{E}(Y) - \hat{\beta}_2 \mathbb{E}(X_2), \frac{Cov(X_2, Y)}{Var(X_2)})$ を得る.

β の推定値　今, $(X_{1,2}, Y_1), \dots, (X_{n,2}, Y_n)$ と n 個の無作為標本が与えられ, その**実現値** $(x_{1,2}, y_1), \dots, (x_{n,2}, y_n)$ が判明している下で, MSE を**標本平均に置き換え**, 以下の二乗誤差の標本平均（≈ 標本二乗残差（後述））の minimizer として OLS 推定値（× 推定量）(b_1, b_2) を決定したい.

$$\bar{Loss} = \frac{1}{n} S(b_{1,0}, b_{2,0}) = \frac{1}{n} \sum_{i=1}^{n} e_{i,0}^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - b_{1,0} - b_{2,0} X_{i,2})^2$$

この結果, OLS 推定量 $(b_1, b_2) = (\bar{Y} - b_2 \bar{X}_2, \frac{s_{X_2,Y}}{s_{X_2}^2})$ を得る. さらに実データに置き換えれば, $(b_1, b_2) = (\bar{y} - b_2 \bar{x}_2, \frac{s_{x_2,y}}{s_{x_2}^2})$ の OLS 推定値が得られる.

y_i の予測値　次に, 無作為標本の実現値が与えられ, OLS 推定値を用いて導かれる内生変数の予測値に関連する性質を考える. 予測値 $\hat{y}_i = b_1 + b_2 x_{i,2}$, 残差 $e = y_i - \hat{y}_i$ とする. このとき, $\bar{e} = 0$, $s_{e,x_2} = 0$, $\bar{\hat{y}} = \bar{y}$. e は ϵ の実現値と見なせるため, $\mathbb{E}[\epsilon] = 0$, $\mathbb{E}[X_2 \epsilon] = 0$, $\mathbb{E}[h(X_2)\epsilon] = 0$ に対応した結果であることも分かる. 残差の定義が, 決定係数の定義の解釈に有用であることも留意.

標本外の y_i の予測　ここからは, 同様の線形単回帰モデルにおいて, n 組の無作為標本の実現値が与えられ, ここから構成された OLS 推定値 (b_1, b_2) を用いた, **標本外**の内生変数の予測を目標として議論を進める. $\{X_i\}_{i=1}^n = \{x_i\}_{i=1}^n$ で実現値が与えられており, 単回帰モデルに従って $\{Y_i\}_{i=1}^n = \{\beta_1 + \beta_2 x_{2,i} + \epsilon_i\}_{i=1}^n$ が生成されるものの, 実現値は未だ与えられていない（$\because \exists \epsilon_i$）場合を考える. この時の β の推定量（× 推定値）は $(b_1, b_2) = (\bar{Y} - b_2 \bar{x}_2, \frac{s_{x_2,Y}}{s_{x_2}^2})$ となり, ϵ の

関数なので確率関数として表せて, この期待値をとれば OLS 推定量の不偏性を示すことが出来る．更に, 同一母集団同時分布に従う, 標本外の確率変数 $(X_{i,2}^{OOS}, Y_i^{OOS})$ を考える．我々のここでの目標は $X_{i,2}^{OOS} = x_i^{OOS}$ として実現値を与えられたときの Y_i^{OOS} を予測することにある．$\mathbb{E}[\hat{Y}_i^{OOS}|X_{i,2}^{OOS} = x_{i,2}^{OOS}] = \mathbb{E}[b_1 + b_2 x_i^{OOS}|X_{i,2}^{OOS} = x_{i,2}^{OOS}] = \beta_1 + \beta_2 x_i^{OOS} = \mathbb{E}[Y_i^{OOS}|X_{i,2}^{OOS} = x_{i,2}^{OOS}]$（$\because$ OLSE の不偏性, $\mathbb{E}[\epsilon|X] = 0$）なので, **この OLS 予測それ自体も不偏性を持つ**と言える．

線形射影　先ほど扱った期待値表記の OLS method は, 線形射影係数の識別にも利用される．これはデータの最適な線形近似を与え, 実際に条件付き期待値が線形で無かったとしても成り立つ. OLS の理論を補強できるものとして考えられる．

誘導型と構造型　回帰モデルにおいて, β は必ずしも $\mathbf{X}$ の因果効果を意味する必要はない．まず, 因果効果として解釈できない場合, すなわち偶然変数同士が相関しており, $\mathbb{E}(Y|\mathbf{X}) = \mu(\mathbf{X})$ を満たすような関係を示す**誘導型モデル**を考える．この場合, 推定値 b は, $\mathbf{X}$ の変化による Yへの影響の程度を捉えるが, それは単なる統計的な共変動（$\times$ 実際の因果関係）を意味するに過ぎない．したがって、$\mathbf{X}$ が他の変数と相関していても（Ex. 内生変数）**モデル推定自体は可能**, その結果をもとに**予測も可能**. 今までのモデルは予測に注目しており, 誘導型モデルだったと言える．一方, **構造型モデル**として回帰モデルを捉える場合, β の解釈は異なる．ここでは, β は $\mathbf{X}$ の Yに与える**直接的因果効果**の大きさ, 即ち『X_i が 1 単位増加したとき, 他全て（の変数）を一定に Yがどれだけ変化するか』として解釈される．つまり, 因果効果を正確に捉えることが目的であるため, 因果推論を成立させるための前提条件（Ex. $\mathbf{X}$ の外生性）が重要となる（$\Rightarrow$ IV 法）．

因果推論のための単回帰モデル・古典的仮定　では, 実際に β_2 が因果効果として考えるための仮定とは何だろうか. これ即ち**単回帰モデルの古典的仮定**である．

(**A1**)$\mathbb{E}(\epsilon_i|X_{i,2}) = 0$
(**A2**)$(X_{i,2}, Y_i)$　for　$i = 1, \dots, n$　は i.i.d.
(**A3**)$(X_{i,2}, \epsilon_i)$は無限ではない 4 次のモーメントを持つ
(**A4**)均一分散 : $Var(\epsilon_i|X_{i,2}) = \sigma_\epsilon^2$
(**A5**)条件付き正規性 : $\epsilon_i|X_{i,2} \sim \mathcal{N}(0, \sigma_\epsilon^2)$

以上を満たすのが, 因果推論のための構造型単回帰モデルとして標準的なものである．このモデルでは, β_2 は $X_{i,2}$ の変化による**因果効果**として定義されることに留意せよ．当然見せかけの因果関係や逆の因果関係は事前に排除されている．ここで, $X_{i,2}$ がランダムに割り当てられている（少なくともそのように見える）ならば, 構造モデルに必要な, 説明変数の誤差項との無相関, すなわち**外生性**を示唆する **(A1)** が満たされる．数学的には, (A1)$\Rightarrow Cov(\epsilon_i, X_{i,2}) = 0$ であり, 対偶をとれば, $Cov(\epsilon_i, X_{i,2}) \neq 0 \Rightarrow$ (A1) が崩れると表現可能. (A1) は, b_2 が β_2 の不偏推定量であることを確約している．説明変数の漏れにより, b_2 がもはや β_2 の一致推定量でなくなる（不偏推定量でもない）, つ

まり**欠落変数バイアス** (後述) による (A1) の崩れは, まさしく操作変数法に繋がる議論である.

b_2 の検定　b_2 の両側検定 $H_0 : \beta_2 = \beta_{2,0}$ について検討する. 先述の通り $b_2 = \frac{s_{X,Y}}{s_X^2}$ が β_2 の推定量である. 説明変数 X の母集団分布についての情報が未知な状況を考えると, t 検定統計量は $T = \frac{b_2-\beta_{2,0}}{\sqrt{\hat{Var}(b_2)}} = \frac{b_2-\beta_{2,0}}{SE(b_2)}$ である. $\epsilon|X \sim \mathcal{N}(0,\epsilon^2)$ を仮定した finite sample のケースにおいては, b_2 を分解してやり, σ_X^2 が未知だったケースのように変形を行えば $T \sim t(n-2)$ であることが分かる. 次に, 漸近理論から考察を行う. 今回は誤差項の正規性の仮定は置かず, $(X_i - \bar{X})\epsilon_i$ に CLT が適応可能との仮定 (と中級計量にはあるが ...?), 言い換えればこの標準化標本平均に関する**漸近正規性**を仮定する. 仮定のもとで, 標本数が十分に大きければ, $T \xrightarrow{d} \mathcal{N}(0,1)$**(未確認, 中級計量 Ch3 の後半参照)** が成立し, ここでも t 統計量の分布は標準正規分布で良く近似できることがわかった. 最後に信頼区間について考えよう. 再び誤差項に正規性の仮定をおけば, 分散が未知の場合でも $T \sim t(n-2)$ であったことを利用して, $100(1-\alpha)$ パーセント信頼区間 $[b_2 - t_{1-\alpha/2}(n-2)\widehat{SE}(b_2), b_2 + t_{1-\alpha/2}(n-2)\widehat{SE}(b_2)]$ が構成できる.

2.4　重回帰モデルの導入

線形重回帰モデルも単回帰と同様に, 実際に条件付き期待値が $\mathbb{E}[Y|\mathbf{X}] = \mathbf{X}\beta$ である（＝変数同士が実際に線形関係に従っている）と仮定する. 議論の筋は単回帰のそれとおよそ同じであり, 違いは説明変数が複数ある事, それに伴い議論に行列を利用することである.

重回帰モデルの古典的仮定　回帰係数ベクトル β が因果効果であることを以下が保証する.

(**A1**)$\mathbb{E}(\epsilon_i|\mathbf{x}_i) = 0$

(**A2**)$(\mathbf{x}_i, Y_i)$　for　$i = 1, \dots, n$　は i.i.d.

(**A3**)$(\mathbf{x}_i, \epsilon_i)$は無限ではない 4 次のモーメントを持つ

(**A4**)フルランク (識別) : $rank(\mathbf{X}) = K$

(**A5**)均一分散 : $Var(\epsilon_i|\mathbf{x}_i) = \mathbb{E}(\epsilon_i^2|\mathbf{x}_i) = \sigma^2$

(**A6**)条件付き正規性 : $\epsilon_i|\mathbf{x}_i \sim \mathcal{N}(0,\sigma^2)$

(A1) は LIE より $\mathbb{E}[\epsilon_i] = 0, Cov(\epsilon_i, x_{i,k}) = 0 \forall k = 1, \dots, K$ であり, 説明変数の外生性を保証する. (A1),(A2) によって, 以下の関係が成立する.

$$\mathbf{E}[\epsilon|\mathbf{X}] = \begin{pmatrix} \mathbb{E}[\epsilon_1|\mathbf{X}] \\ \vdots \\ \mathbb{E}[\epsilon_n|\mathbf{X}] \end{pmatrix} = \begin{pmatrix} \mathbb{E}[\epsilon_1|\mathbf{x}_1] \\ \vdots \\ \mathbb{E}[\epsilon_n|\mathbf{x}_n] \end{pmatrix} = \mathbf{0}$$

$$\mathbf{E}[\mathbf{Y}|\mathbf{X}] = \mathbf{X}\beta + \mathbf{E}[\epsilon|\mathbf{X}] = \mathbf{X}\beta$$

$$Cov(\epsilon_i, \epsilon_j|\mathbf{X}) = \mathbb{E}(\epsilon_i\epsilon_j|\mathbf{X}) = \mathbb{E}(\epsilon_i\epsilon_j|\mathbf{x}_i, \mathbf{x}_j) = \mathbb{E}(\epsilon_i|\mathbf{x}_i)\mathbb{E}(\epsilon_j|\mathbf{x}_j) = 0$$

(A3) は漸近理論のために必要な仮定であり, (A4) は説明変数間に完全な線形関係（多重共線性）がないことを保証しており, $rank(\mathbf{X}^T\mathbf{X}) = rank(\mathbf{X}) = K$ も成立する. (A4) が崩れた場合, 回帰係数は無限個解を持つ (不定) ため, 識別が不可能となる. 特にダミー変数について, この状況は**ダミー変数の罠 (dummy variable trap)** と呼ばれる. (A1), (A2), (A5) より, $\mathbf{Var}(\epsilon|\mathbf{X}) = \mathbf{E}[\epsilon\epsilon^T|\mathbf{X}] = \sigma^2\mathbf{I}$. さらに全分散の法則より, $\mathbf{Var}(\epsilon) = \mathbf{E}[\mathbf{Var}(\epsilon|\mathbf{X})] + \mathbf{Var}(\mathbf{E}[\epsilon|\mathbf{X}]) = \sigma^2\mathbf{I}$, LIE より $\mathbf{E}[\epsilon] = \mathbf{0}$ を得て直接計算しても確認可能である. (A1), (A2), (A5), (A6) から, $\epsilon|\mathbf{X} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$.

βの推定　今, $(\mathbf{x}_1, Y_1), \ldots, (\mathbf{x}_n, Y_n)$ と n 個の無作為標本が与えられ, その**実現値** $(\mathbb{x}_1, y_1), \ldots, (\mathbb{x}_n, y_n)$ が判明している下で, 以下の残差平方和の minimizer として OLS 推定値 $\mathbf{b}$ を決定する.

$$S(\mathbf{b}_0) = \sum_{i=1}^{n} e_{i,0}^2 = \mathbf{e}_0^T\mathbf{e}_0 = (\mathbf{y} - \mathbb{X}\mathbf{b}_0)^T(\mathbf{y} - \mathbb{X}\mathbf{b}_0)$$

以上より OLS 推定値 $\mathbf{b} = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbf{y}$ を得られる. 一般に, $\hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ が OLS 推定量.

補足:OLSE の導出　行列計算の練習に OLS 推定値の導出を確認しよう. $S(\mathbf{b}_0) = \sum_{i=1}^{n} e_{i,0}^2 = \mathbf{e}_0^T\mathbf{e}_0 = (\mathbf{y} - \mathbb{X}\mathbf{b}_0)^T(\mathbf{y} - \mathbb{X}\mathbf{b}_0) = (\mathbf{y}^T - \mathbf{b}_0^T\mathbb{X}^T)(\mathbf{y} - \mathbb{X}\mathbf{b}_0) = \mathbf{y}^T\mathbf{y} - \mathbf{b}_0^T\mathbb{X}^T\mathbf{y} - \mathbf{y}^T\mathbb{X}\mathbf{b}_0 + \mathbf{b}_0^T\mathbb{X}^T\mathbb{X}\mathbf{b}_0 = \mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\mathbb{X}\mathbf{b}_0 + \mathbf{b}_0^T\mathbb{X}^T\mathbb{X}\mathbf{b}_0$, 最後の変形は第 $3, 4$ 項がスカラーであることによる. 正規方程式は以下のように変形できる. $\mathbf{0} = \frac{\partial S(\mathbf{b}_0)}{\partial \mathbf{b}_0}|_{\mathbf{b}_0=\hat{\mathbf{b}}_*} = -2\frac{\partial S(\mathbf{b}_0)}{\partial \mathbf{b}_0}\mathbf{y}^T\mathbb{X}\mathbf{b}_0|_{\mathbf{b}_0=\hat{\mathbf{b}}_*} + \frac{\partial S(\mathbf{b}_0)}{\partial \mathbf{b}_0}\mathbf{b}_0^T\mathbb{X}^T\mathbb{X}\mathbf{b}_0|_{\mathbf{b}_0=\hat{\mathbf{b}}_*} = -2(\mathbf{y}^T\mathbb{X})^T + (\mathbb{X}^T\mathbb{X} + (\mathbb{X}^T\mathbb{X})^T)\hat{\mathbf{b}}_* = -2\mathbb{X}^T\mathbf{y} + 2\mathbb{X}^T\mathbb{X}\hat{\mathbf{b}}_*$. $\mathbb{X}^T\mathbf{y} = \mathbb{X}^T\mathbb{X}\hat{\mathbf{b}}_*$ から, $(\mathbb{X}^T\mathbb{X})^{-1}(\mathbb{X}^T\mathbb{X})\hat{\mathbf{b}}_* = \hat{\mathbf{b}}_* = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbf{y}$.

$\mathbf{y}$ の予測値　次に, 無作為標本の実現値が与えられ, OLS 推定値を用いて導かれる内生変数の予測値に関連する性質を考える. 予測値 $\hat{y}_i = \mathbb{x}_i^T\mathbf{b}$, 残差 $e = \mathbf{y} - \hat{\mathbf{y}}$ とする. このとき, 正規方程式 $\mathbb{X}^T(\mathbb{y} - \mathbb{X}\mathbf{b}) = \mathbf{0}$ より $\mathbb{X}^T\mathbf{e} = \mathbf{0}$ である. 特に, 一般的な重回帰モデルでは $\mathbf{X}$ の一列目が 1 であることから, $\frac{1}{n}\sum_{i=1}^{n} e_i = \bar{e} = 0$, $s_{e,x_k} = 0 \quad \forall k = 2, \ldots, K$, $\bar{y} = \bar{\mathbb{x}}^T\mathbf{b}$($\bar{\mathbb{x}}$ は説明変数の標本平均ベクトル), $\bar{\hat{y}} = \bar{y}$ が成り立つ. 決定係数についても単回帰と同様で, $R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS} = 1 - \frac{\mathbf{e}^T\mathbf{e}}{\mathbf{y}^T\mathbf{y} - n\bar{y}^2} = (\hat{corr}(y_i, \hat{y}_i))^2$. 説明変数が増えれば R^2 は必ず上昇する性質がある. この問題を解決する為に**調整済み決定係数**も存在し, 通常の R^2 とは $\bar{R}^2 = 1 - \frac{n-1}{n-K}(1 - R^2)$ の関係が成立する. ただ決定係数は予測のデータへの当てはまりを示しているに過ぎないため, この数値が低いからと言って, 特に因果推論に用いる場合には必ずしもモデルが悪いわけではないことに注意が必要である. また, 予測に関しても線形射影としての OLS の有用性を考えれば, データが非線形に生成されるケースには決定係数は悪くな

れど依然線形近似としての意義はある．このような問題からか，末石計量ではもはや決定係数に関する記載は一切ない．

3 回帰モデルの性質

3.1 Finite Sample の性質

有限の標本について成立する性質を検討する．

OLS 推定量の性質　まず，不偏性について考えよう．

ガウスマルコフの定理　ガウスマルコフの定理 (Gauss-Markov Theorem): 回帰モデルの古典的仮定のもとでは，OLS 推定量 $\mathbf{b}$ は線形不偏推定量のうちで最も効率的である．ここでまず効率性の定義を確認しよう．二つの不偏推定量 $\hat{\theta}$ と $\tilde{\theta}$ を考えたとき，$Var(\mathbf{a}^T\tilde{\theta}) - Var(\mathbf{a}^T\hat{\theta}) = \mathbf{a}^T[Var(\tilde{\theta}) - Var(\hat{\theta})]\mathbf{a} \geq 0 \quad \forall \mathbf{a} \neq \mathbf{0}$ ならば，$\hat{\theta}$ が効率的という．つまり共分散行列の差が非負値定符号行列である．共分散行列の差は対称行列より，効率的であることは，そのすべての主小行列式（Principal Minors, PM）が非負であることと同値．また，全ての対角成分，即ち各々の分散が他の推定量に比べ最低限大きくないことが必要条件である．1 次元の推定量の場合，効率的であることは分散が大きくないことと同値で，非負定符号行列を用いた上記の定義はその一般化である．ガウスマルコフの定理は，この効率性の定義を利用して，全ての不偏推定量 $\mathbf{b}_0$ に関して，$\mathbf{Var}[\mathbf{b}_0|\mathbf{X}] - \mathbf{Var}[\mathbf{b}|\mathbf{X}]$ が非負値定符号行列であることを主張していると言い換えられる．

4 プログラム評価

我々が関心を持つ対象が，X が一単位変化したときの Y への**因果効果，処置効果**だとする．このセクションではこの効果を分析するプログラム評価の基礎的な概念をレビューする．個人 i が処置群に含まれるなら $X_i = 1$，対照群に含まれるならば $X_i = 0$ をとる確率変数を考える．個人 i の処置効果は，$te_i = Y_{i,1} - Y_{i,0} = Y_i(X_i = 1) - Y_i(X_i = 0)$ だが，このうちどちらかは反実仮想であり，従って観察不可能である．そこで処置の**平均処置効果，ATE** を考えると，これは $\mathbb{E}[Y_{i,1}] - \mathbb{E}[Y_{i,0}]$ で与えられる．また，処置群への平均処置効果 (ATET) は $\mathbb{E}[Y_{i,1}|X_i = 1] - \mathbb{E}[Y_{i,0}|X_i = 1]$，対照群への平均処置効果 (ATENT) は $\mathbb{E}[Y_{i,1}|X_i = 0] - \mathbb{E}[Y_{i,0}|X_i = 0]$ である．現状ではどれも観察不可能な項が含まれることに注意せよ．現実に観測でき，かつ一見因果効果を捉えているように思えるのは $\mathbb{E}[Y_{i,1}|X_i = 1] - \mathbb{E}[Y_{i,0}|X_i = 0]$ だろう．しかし，以下の式変形で問題点が明らかになる．

$$\mathbb{E}[Y_{i,1}|X_i=1]-\mathbb{E}[Y_{i,0}|X_i=0]=ATE+\underbrace{\mathbb{E}[Y_{i,0}|X_i=1]-\mathbb{E}[Y_{i,0}|X_i=0]}_{\text{selection bias}}+\underbrace{(1-\pi)(ATET-ATENT)}_{\text{heterogeneous selection bias}}$$

ここで, π は処置群に含まれる確率である. 因果効果を議論する際, 処置群と対照群がランダムに割り当てられていない限りは, セレクションバイアスに注意する必要性が明らかとなった.

ランダム比較実験, RCT　では, どのような条件の下で $\mathbb{E}[Y_{i,1}|X_i=1]-\mathbb{E}[Y_{i,0}|X_i=0]=ATE$ は保証されるだろうか? この十分条件を満たすのが既に幾度か言及している**ランダム比較実験, RCT** である. 処置群と対照群がランダムに割り当てられているこの状況下では, 上記の式が $\mathbb{E}[Y_{i,1}|X_i=1]=\mathbb{E}[Y_{i,1}|X_i=0]=\mathbb{E}[Y_{i,1}]$, $\mathbb{E}[Y_{i,0}|X_i=1]=\mathbb{E}[Y_{i,0}|X_i=0]=\mathbb{E}[Y_{i,0}]$ となることで満たされる. また, 各人で同一の固定処置効果を仮定した場合, 単回帰の傾きの OLSE が ATE に整合することも知られており, この結果を補強している. 詳細は次の通り. $Y_{i,1}=\tau+Y_{i,0}$ より, $Y_i=Y_{i,0}+(Y_{i,0}-Y_{i,1})X_i=Y_{i,0}+\tau X_i$ で, $Y_{i,0}=\mathbb{E}[Y_{i,0}]+\epsilon_{i,0}$ とすれば $Y_i=\mathbb{E}[Y_{i,0}]+\tau X_i+\epsilon_{i,0}=\beta_1+\beta_2 X_i+\epsilon_i$ と単回帰モデルの形になる. (A1) が成立している, 即ち $\mathbb{E}[\epsilon_i|X_i]=\mathbb{E}[\epsilon_{i,0}|X_i]=0$ が成り立つならば, すなわち RCT の仮定の下では, OLS の β_2 への不偏性は保証されている. 特にこの場合 τ の推定値が単回帰モデルの形でも正確に因果効果を意味することが分かるだろう. 個人の処置効果が不均一でもこの議論は行える. $Y_{i,1}=\mathbb{E}[Y_{i,1}]+\epsilon_{i,1}$ とすると, $te_i=\tau_{ATE}+[\epsilon_{i,1}-\epsilon_{i,0}]$. 同様に単回帰モデルの形にすると, $\beta_2=\tau_{ATE}$, $\epsilon_i=\epsilon_{i,0}+[\epsilon_{i,1}-\epsilon_{i,0}]X_i$ となり, RCT の仮定の下では $\mathbb{E}[\epsilon_i|X_i]=0$ が成立する. しかし不均一分散であることに注意. 他にも, 複数の処置を入れた場合や, 処置が連続的の線形モデル (限界効果が一定) などの拡張がある.

5　コントロール変数

我々の関心は, ある変化によりもたらされた帰結＝**因果効果の推定**にあるとする. 先述の通り, 自然実験で得られたデータとは違い, 観測データには様々なノイズが介在しうる. 因果推論の文脈における回帰分析の主要な役割は, 観測データからノイズを除去して因果効果のみを取り出すことにあるのだ. その目的のためには, すべての変数を平等に扱う必要はない. ここで導入されるのが**コントロール変数**と **代理変数**の概念である. コントロール変数の定義としては以下で挙げるようなものがある. 1: 回帰に含まれたとき，誤差項を興味のある説明変数と無相関にするもの. 2: 値を一定に保つと，興味のある説明変数が「あたかも」無作為に割り当てられている状態を作り出せるもの. 3: 値が同じ個人間では，興味のある説明変数が, 省略された決定因子と相関しないようなもの. コントロール変数をモデルに導入することによって, 後述する欠落変数バイアスの発生を防ぎ, 興味のある説明変数のパラメータを正確に推定可能にしている. ある要素が観測不可能である時に利用されるのが, その要素と強相関がある**代理変数 (proxy variables)** である. コントロール変数の文脈での代理変数は, 観測不可能なコントロール対象がある時に, 対象と相関するが, それ自体は因果

効果を持たないもの, である.

欠落変数と因果推論　古典的仮定を満たす線形モデル $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \tilde{\epsilon}$ に従って生成される観測データから, X_2 による因果効果 β_2 を測定することを目標として考える. Y を $1, X_2$ に回帰してもその測定は可能なのだろうか? $\mathbb{E}[Y|X_2] = \beta_1 + \beta_2 X_2 + \beta_3 \mathbb{E}[X_3|X_2] + \mathbb{E}[\tilde{\epsilon}|X_2]$ で, 真の回帰モデルにおける (A1) より $\mathbb{E}[\tilde{\epsilon}|X_2] = \mathbb{E}[\mathbb{E}[\tilde{\epsilon}|X_2, X_3]|X_2] = 0$. ここで, $X_3 = \delta_1 + \delta_2 X_2 + \eta$, $\mathbb{E}[\eta|X_2] = 0$, $\delta_2 \neq 0$ と仮定すれば, $\mathbb{E}[Y|X_2] = \beta_1 + \beta_2 X_2 + \beta_3 \mathbb{E}[\delta_1 + \delta_2 X_2 + \eta|X_2] = (\beta_1 + \beta_3\delta_1) + (\beta_2 + \beta_3\delta_2)X_2$ である. $Y = \mathbb{E}[Y|X_2] + \epsilon = (\beta_1 + \beta_3\delta_1) + (\beta_2 + \beta_3\delta_2)X_2 + \epsilon$ から, $\mathbb{E}[\epsilon|X_2] = 0$ であるため, (A1) はこの仮定下では満たせている. しかしここで推定の対象である, X_2 の切片である $\beta_2 + \beta_3\delta_2$ は真の β_2 と異なるため, モデルは因果効果の測定には利用できない.

Ex. 出生時体重の因果推論　幼児の健康状態を決定する重要な要素として出生児体重がある. 今回はこの数値 $weight$ に関する因果推論を行うことを目標として定めよう. $weight$ を減少させる因果効果を持つ要素の一つとして, 母親の妊娠時喫煙が考えられる. 今回は, 喫煙していれば 1 を, してなければ 0 をとるダミー変数 $smoke$ を導入した, 単回帰モデル $weight = \beta_1 + \beta_2 smoke + \epsilon$ について考える. このモデルは因果効果を正しく推定できるのだろうか? **解答**　他に因果効果をもつ要素として, 所得などが考えられる. これらの効果をコントロールしていない限りは不可能. ただし体重の予測は依然可能.

欠落変数バイアス　再度, 古典的仮定を満たす線形モデルに従って生成される観測データ $Y_i = \beta_1 + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \tilde{\epsilon}_i$, $i = 1, \ldots, n$ から, X_2 による因果効果を測定することを目標として考える. Y_i を $1, X_{i,2}$ に回帰したときでも, $Y_i = \beta_1 + \beta_2 X_{i,2} + \epsilon_i$ with $\epsilon_i = \beta_3 X_{i,3} + \tilde{\epsilon}_i$ で, 誤差項に欠落変数の影響を全て格納し, 正確に因果効果を測れるような状況を作り出すために, 如何なる条件があるのかをより一般的に考えてみよう. OLS 推定量の公式を適用すると, $b_2' = \frac{s_{X_2,Y}}{s^2_{X_2}} = \frac{\frac{1}{n-1}\sum (X_{i,2} - \bar{X_2})(Y_i - \bar{Y})}{\frac{1}{n-1}\sum (X_{i,2} - \bar{X_2})^2} = \beta_2 + \frac{\frac{1}{n-1}\sum (X_{i,2} - \bar{X_2})(\epsilon_i - \bar{\epsilon_i})}{\frac{1}{n-1}\sum (X_{i,2} - \bar{X_2})^2} = \beta_2 + \frac{\frac{1}{n-1}\sum (X_{i,2} - \bar{X_2})\epsilon_i}{\frac{1}{n-1}\sum (X_{i,2} - \bar{X_2})^2}$ ($\because Y_i - \bar{Y} = (\beta_1 + \beta_2 X_{i,2} + \epsilon_i) - (\beta_1 + \beta_2 \bar{X_2} + \frac{1}{n}\sum \epsilon_i) = \beta_2 (X_{i,2} - \bar{X_2}) + (\epsilon_i - \bar{\epsilon_i}), \sum (X_{i,2} - \bar{X_2})\bar{\epsilon} = \bar{\epsilon} \sum (X_{i,2} - \bar{X_2}) = 0$). ここで, n が十分大きい場合, LLN より $\frac{1}{n-1}\sum (X_{i,2} - \bar{X_2})^2 \xrightarrow{p} Var(X_2)$, $\frac{1}{n-1}\sum (X_{i,2} - \bar{X_2})(\epsilon_i - \bar{\epsilon_i}) \xrightarrow{p} Cov(X_2, \epsilon)$ が成立する. このため, $b_2' \xrightarrow{p} \beta_2 + \frac{Cov(X_2,\epsilon)}{Var(X_2)}$ である. ここで, (A1)$\mathbb{E}[\epsilon|X_2] = 0 \implies Cov(X_2, \epsilon) = 0$ は一致性 $b_2' \xrightarrow{p} \beta_2$ の十分条件である. また, 不偏性についても, (A1) を仮定するならば, X_2 についての条件付期待値をとって LIE を利用すれば確認可能. $\epsilon_i = \beta_3 X_{i,3} + \tilde{\epsilon}_i$ を利用して更に考察を進めてみよう. $Cov(X_2, \epsilon) = Cov(X_2, \beta_3 X_3 + \tilde{\epsilon}) = \beta_3 Cov(X_2, X_3) + Cov(X_2, \tilde{\epsilon})$. ここで真の構造型回帰モデルについての (A1) より $\mathbb{E}[\tilde{\epsilon}|X_2, X_3] = 0$ が成立しているため, LIE より $Cov(X_2, \tilde{\epsilon}) = \mathbb{E}[(X_2 - \bar{X_2})(\tilde{\epsilon} - \bar{\tilde{\epsilon}})] = \mathbb{E}[\mathbb{E}[(X_2 - \bar{X_2})(\tilde{\epsilon} - \bar{\tilde{\epsilon}})]|X_2, X_3] = \mathbb{E}[(X_2 - \bar{X_2})\mathbb{E}[\tilde{\epsilon} - \bar{\tilde{\epsilon}}|X_2, X_3]] = 0$. ゆえに $Cov(X_2, \epsilon) = \beta_3 Cov(X_2, X_3)$ であり, 以下のように書ける.

$$b_2' \xrightarrow{p} \beta_2 + \underbrace{\beta_3 \frac{Cov(X_2, X_3)}{\sigma^2_{X_2}}}_{\text{omitted variable bias}}$$

RCT の仮定下では X_2 はランダムに割り当てられているため, b_1' は一致性と不偏性を満たすが, $\beta_3 \neq 0$ nor $Cov(X_2, X_3) \neq 0$ の時には一致性を持たず, このズレのことを欠落変数バイアス (omitted variable bias) と呼ぶ. 補足できていない欠落変数の因果効果が, 相関する変数 X_2 を経由して被説明変数に影響を与えており, 推定のズレの方向は相関に一致する. この時, 先述の通り $Cov(X_2, \epsilon) \neq 0 \implies \mathbb{E}[\epsilon|X_2] \neq 0$ で (A1) が崩れていることを確認せよ. この時, 説明変数に**内生性**があるという.

コントロール変数 古典的仮定を満たす線形モデル $Y_i = \beta_1 + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \epsilon_i$ に従って生成される観測データから, X_2 による因果効果を測定することを目標として考える. Y_i を $1, X_{i,2}$ に回帰すると, $\beta_3 \neq 0$ nor $Cov(X_2, X_3) \neq 0$ の場合は, 欠落変数バイアスから正確に因果効果を測定できない. コントロール変数として $X_{i,3}$ を入れてやり, $1, X_{i,2}, X_{i,3}$ に回帰すれば, $E[\epsilon_i|X_{i,2}, X_{i,3}] = 0$ で b_2, b_3 は共に一致推定量となり, 正しく因果効果を測定できる.

冗長なコントロール変数 コントロール変数を不用意に追加することは, 不偏性には影響を及ぼさずとも, 推定量の分散を大きくし, 精度を悪化させることが知られている. これを詳しく見てみよう. $Y = \beta_1 + \beta_2 X_2 + \tilde{\epsilon}$ を, 余分に説明変数を付け加えた $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$ で誤って推定する場合を考える. $\beta_3 = 0$ とすると, 欠落変数バイアスは発生せず, $E[\epsilon|X_2, X_3] = 0$ が成立していれば誤ったモデルでも真のパラメータとの一致性が保たれる. 大標本においては, $b_2 \overset{a}{\sim} \mathcal{N}(\beta_2, \sigma^2_{b_2})$, where $\sigma^2_{b_2} = \frac{1}{n}(\frac{1}{1-\rho^2_{X_2,X_3}})\frac{\sigma^2_\epsilon}{\sigma^2_{X_2}}$ であることが知られている. この結果より, 冗長な変数ともとの説明変数との間に強い相関 ρ_{X_2,X_3} があればあるほど分散が大きくなり, 推定は悪化することが分かった.

変数選択 一方で RCT の仮定下で議論をすすめれば, $\rho_{X_2,X_3} = 0$, $Var(\epsilon) > Var(\tilde{\epsilon})$ より, 適切な変数追加ならば, 予測精度を向上させることも確認できる. 変数追加について慎重になるべきケースはまだ存在する. 現在の説明変数の因果効果を媒介する新たな説明変数は, 因果効果を推定できなくなるため入れるべきではない. このようなコントロール変数を**悪いコントロール (bad control)** とよぶ.

議論の一般化 欠落変数の因果推論への影響を一般化して考えよう. 類似の議論は北村の応用ミクロ計量 (2002)-第二講 (`https://www.ier.hit-u.ac.jp/~kitamura/index_j.html`) にて展開されている, さらに言えば元文献は Wooldridge (2002, Ch.4) らしいが. 構造モデルが $\mathbb{E}[y|x_2, \ldots, x_k, q] = \beta_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \gamma q$ with $\gamma \neq 0$ で与えられるとしよう. これを書き直せば, $y = \beta_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \gamma q + \tilde{\epsilon}$ で, $\mathbb{E}[\tilde{\epsilon}|x_2, \ldots, x_k, q] = 0$ より (A1) を満たす. ここで, 観

測できない**潜在変数** q をモデルに直接入れ込むことは不可能であるため, 実際に推定する際のモデルは $y = \beta_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \epsilon$ where $\epsilon = \gamma q + \tilde{\epsilon}$. このモデルでの因果推論が可能, つまり欠落変数バイアスがないのは, q が $\mathbf{x}$ それぞれと無相関である場合に限られる. 仮定のもとで q を $\mathbf{x}$ に回帰すると $q = \delta_1 + \delta_2 x_2 + \cdots + \delta_K x_K + r$ となり, $Cov(x_j, r) = 0 \forall j$. 代入すると, $y = (\beta_1 + \gamma\delta_1) + (\beta_2 + \gamma\delta_2)x_2 + \cdots + (\beta_K + \gamma\delta_K)x_K + \tilde{\epsilon} + \gamma r$, with $Cov(x_j, \tilde{\epsilon} + \gamma r) = 0 \forall j$ を得た. 欠落変数バイアスの部分と同様の議論によって, $b_k \xrightarrow{p} \beta_k + \gamma\delta_k$ で一致性の崩れを確認できる. この場合の欠落変数バイアスを防ぐ方法が**代理変数**の利用だ. この際代理変数が満たすべき条件は以下.

1. 代理変数が構造モデルの中で重複している: z を q の代理変数として, $\mathbb{E}[Y|\mathbf{x}, q, z] = \mathbb{E}[Y|\mathbf{x}, q]$ が成立する. つまり Y の説明に z が無関係であること.

2. q と x_j の相関は z の影響をコントロールするとゼロになる: つまり $q = L[q|1, \mathbf{x}, z] = L[q|1, z] = \delta_1 + \delta_2 z + r$, where $\delta_2 \neq 0, \mathbb{E}[r] = 0, Cov(z, r) = 0$. 同値の表現として $Cov(x_j, r) = 0 \forall j$ も.

上記の条件のもと, 構造モデルは $Y = (\beta_1 + \gamma\delta_1) + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_K x_K + \gamma\delta_2 z + (\gamma r + v)$ となり $\mathbf{x}$ での OLS 推定は一致性を保っている.

コントロール変数の文脈における代理変数 再び構造モデル $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \tilde{\epsilon}$ にもどり, 代理変数の活用法を考える. 古典的仮定を満たす上記の線形モデルに従って生成される観測データから, X_2 による因果効果を測定することを目標として考える. ただし X_3 は観測不能とする. $1, X_2$ のみに回帰すると, 欠落変数バイアスから正確に因果効果を測定できないことは先述の通り. ここで, X_3 と相関する, しかし自身は因果効果を持つ必要のない**代理変数** W をモデルに入れ込み, $Y = \beta_1 + \beta_2 X_2 + \beta_3 W + \epsilon$ と書き換え, 以下代理変数の条件付けを言い換えながらこのコントロールの正統性を主張してみよう.

その 1. 代理変数の条件利用 先述の代理変数の 2 条件を適応させて考えてみよう. 条件 1 の言い換えとして, $Cov(W, \tilde{\epsilon}) = 0$ を, 条件 2 を利用して, $X_3 = L[X_3|1, X_2, W] = L[X_3|1, W] = \delta_1 + \delta_2 W + r$, $\delta_2 \neq 0$ と仮定を課しておく. LIE より $\mathbb{E}[r|X_2, W] = 0$. 代入すると $Y = (\beta_1 + \beta_3\delta_1) + \beta_2 X_2 + \beta_3\delta_2 W + \tilde{\epsilon} + \beta_3 r, \mathbb{E}[\tilde{\epsilon} + \beta_3 r|X_2, W] = 0$ を得る. この時 $b_2 \xrightarrow{p} \beta_2$ で因果効果を一致推定できた.

その 2. 条件付平均独立の仮定 (A1) を以下の**条件付平均独立**に置き換えてもよい: $E[\epsilon|X_2, W] = E[\epsilon|W]$. この仮定が β_2 の因果推論の正確性を保証することを確かめるため, $Y = \beta_1 + \beta_2 X_2 + \beta_3 W + \epsilon$ において $E[\epsilon|X_2, W] = E[\epsilon|W] = \gamma_1 + \gamma_2 W, \gamma_2 \neq 0$ と線形関係で条件付平均独立の仮定を満たすよう表現しよう. 更に, ϵ を X_2, W に回帰して, $\epsilon = E[\epsilon|X_2, W] + v$ とすると, LIE より, $E[v|X_2, W] = 0$ である. モデルを明示的に示すと, $Y = \beta_1 + \beta_2 X_2 + \beta_3 W + \epsilon = \beta_1 + \beta_2 X_2 + \beta_3 W + \gamma_1 + \gamma_2 W + v = (\beta_1 + \gamma_1) + \beta_2 X_2 + (\beta_3 + \gamma_2)W + v$ である. このとき, β_2 は依然因果効果として解釈でき, b_2 は一致推定量であるので回帰分析の目標は果たされた. $\beta_3 + \gamma_2$ はバイアスがあっても, そもそも建付けの時点から, 回帰に入れていない変数から W を経由して被説明変数が変化していることを想定しているために問題ない.

Ex. 教育の賃金への因果効果　賃金への影響を与える要因が, 教育年数, 就業年数, そして能力のみであると考えよう. このとき求めたいのは, $\frac{\partial \mathbb{E}(wage|edu,exper,abil)}{\partial edu}$ で, コントロール変数ベクトル $\mathbf{C} = (exper, abil)$ だが能力は**観測不可能**. 回帰モデルを考えると, 就業年数はコントロール変数としてモデルに入れられる他, 能力 $abil$ に相関する代理変数として IQ もコントロール変数に採用できる. すなわち $\mathbb{E}(wage|edu, exper, IQ)$ を考えればよい.

6 内部妥当性

統計的推測の妥当性は, 内部妥当性と外部妥当性の二つに大別される. 内部妥当性とは, 因果効果に関する統計的推測が調査された母集団と設定に対して有効であることを言う. 一方で外部妥当性は, 推論と結論が, 他の集団や設定にも一般化可能であることを意味する. 今回は内部妥当性に焦点を当てて議論を進めよう. 回帰モデルの古典的が外れた時に内部妥当性が崩れる恐れがあるわけだが, さらに, 推定量の不偏性や一致性への影響と, 標準誤差への影響とそれに伴う予測精度の悪化の二つにグループ分けが可能である. 前者の例としては先述の欠落変数バイアスを始めとして, モデルの誤った関数特定化, 測定誤差, セレクションバイアス, 同時性がある. 後者の例としては, 不均一分散や誤差項の内生性などが挙げられる.

測定誤差　説明変数に測定誤差が生じている場合, (A1) が崩れ因果推論が行えなくなる. 例えば誤差 w が分散 σ_w^2 で i.i.d で $K = 2$ のケースをかんがえれば, $b_2 \xrightarrow{p} \beta_2 - \beta_2 \frac{\sigma_w^2}{Var(\tilde{X}_2)}$ で過少推定となる. この解決法には操作変数法が用いられる.

セレクションバイアス　データの可用性が従属変数の値に関連する選択プロセスによって影響を受ける場合に発生する. 誤差項と回帰変数が相関し (A1) が崩れ, OLS 推定の不偏性が崩れる.

Ex. 教育の賃金への因果効果　研究者が, 就職している大学卒業生のランダム標本を調査し, 教育の収益を推定するため賃金と教育年数の回帰をおこなうことを考える. しかし, ここでのサンプルは仕事についている人に限定されている. 就業するか否かを決定する, 教育, 経験, 能力, 運などは, 稼得能力の決定要因に似通っているため, セレクションバイアスが発生していると疑う必要がある.

同時性　変数が相互に因果効果を与えあう場合に起こる問題で, (A1) が崩れる. たとえば, 生徒と教師の比率はテストのスコアに影響するが, 同時に, 政府の取り組みによりテストスコアが低い学区で教師の雇用が補助されるため, テストスコアも生徒と教師の比率に影響する. なお, この問題も操作変数法で解決可能である.

系列相関, 誤差項の内生性　誤差が観測値全体にわたって相関している場合, 仮定 (A2) は破られる．通常パネルデータまたは時系列データで発生する現象である．観測値全体にわたる誤差の相関によって, OLSE の不偏性や一致性が崩れることはないが, 標準誤差が**望ましい有意水準の信頼区間を生成しない**という意味で不正確.

7　制約と F 検定

先ず (A6) 条件付き正規性を仮定した下での議論を行う．F 値の導出には 2 つの手法がある．一つ目は制約式が成立する場合を帰無仮説と置いたとき制約からどれほどずれがあるのかを調べる方法, 二つ目は決定係数を制約ありモデルと制約なしモデルで比較する方法である．R や Stata で表示される F 値は帰無仮説として定数のみ回帰を考えた際の値で, $F(K-1, n-K)$ に従うことに注意せよ．次に, 正規性を外した場合の漸近分布を探ろう．この時, $\mathbf{b}$ の漸近正規性を利用することで $W = FJ \xrightarrow{d} \chi^2(J)$ であることが知られており, この結果は不均一分散のもとでも成立する.

8　構造推定

経済理論から推定された関数形と, 通常の線形モデルを比較しどちらが適切であるかを考察しよう．通常の回帰モデルから外れた概念として, ダミー変数と対数変換, そして多項回帰モデルを本節では導入する．ダミー変数とは, $0, 1$ のどちらか二値を取る変数である．例えば季節ダミーはコントロール変数かつダミー変数である．序数的な意味のみを持つ量的変数に関して気を付けなければいけないのは, それぞれの水準を個別にダミー変数として設定しなければならないことである．というのも, 一つの説明変数で表現を行うことは, 暗黙に一貫した基数的な被説明変数への影響を仮定することだからである．今までは説明変数に関して線形なモデルを考えてきたが, 最初に述べた通り, 我々はパラメータに関して線形であるモデルのヴァリエーションを考慮したい．例としては以下のような変形が挙げられる.

$$\textbf{LoglinearModel} \quad \ln(Y) = \beta_1 + \beta_2 X_2 + \epsilon \tag{5}$$

$$\textbf{LinearlogModel} \quad Y = \beta_1 + \beta_2 \ln(X_2) + \epsilon \tag{6}$$

$$\textbf{LoglogModel} \quad \ln(Y) = \beta_1 + \beta_2 \ln(X_2) + \epsilon \tag{7}$$

(5) ∼ (7) 式については良く知られた通り, 対数を取っている部分は変化率を示している.

$$\textbf{DummyVariable} \quad Y = \beta_1 + \beta_2 X_2 + \beta_3 D + \epsilon \tag{8}$$

$$\textbf{InteractionTerm} \quad Y = \beta_1 + \beta_2 X_2 + \beta_3 X_2 D + \epsilon \tag{9}$$

$$\textbf{Dummy} + \textbf{Interaction} \quad Y = \beta_1 + \beta_2 X_2 + \beta_3 D + \beta_4 X_2 D + \epsilon \tag{10}$$

(8) は切片に, (9) は傾きにダミーの影響を加えるモデルで, 両者を合わせたのが (10) である. 次に**多項回帰モデル (polynomial regression model)**$y_i = \beta_1 + \beta_2 x_{i,2} + \beta_3 x_{i,2}^2 + \cdots + \beta_r x_{i,2}^{r-1} + \epsilon_i$ を考えてみよう. t 検定で r の値を決定し, その後に帰無仮説 $H_0 : \beta_3 = 0, \ldots, \beta_r = 0$ に基づく F 検定で関数形を決定する. これは逐次仮説検定 (sequential hypothesis testing) と呼ばれる.

9 不均一分散

一般化回帰モデルについて考える. 先述の通り, この場合推定量の不偏性は保たれるが, 標準誤差が**望ましい有意水準の信頼区間を生成しない**点で内的妥当性に打撃が与えられる. 解決法としては GLSE による推定と, 頑健標準誤差の利用がある. 一般化回帰モデルの仮定は以下の通り.

(**GLS1**)$\mathbb{E}(\epsilon|\mathbf{X}) = \mathbf{0}$

(**GLS2**)$\mathbf{Var}(\epsilon|\mathbf{X}) = \mathbf{E}[\epsilon\epsilon^T|\mathbf{X}] = \sigma^2\Omega(\mathbf{X}) = \sigma^2\Omega$

(**GLS3**)$(\mathbf{x}_i, \epsilon_i)$は無限ではない 4 次のモーメントを持つ

(**GLS4**)フルランク (識別) : $rank(\mathbf{X}) = K$

(GLS1) が (A1), (A2) のもとでも成り立つことは **2.4** で確認済み. 均一分散と正規性の仮定が外れていることを確認しよう. i.i.d. だが条件付不均一分散をもつデータと, 均一分散だが系列相関を持つデータに GLS は適応できるが, 今回は前者のケースに照準を合わせる. 先ず不偏性を確認しておこう. $\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \beta + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\epsilon$ で, (GLS1) より $\mathbf{E}[\mathbf{b}|\mathbf{X}] = \beta + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{E}[\epsilon|\mathbf{X}] = \beta$ で LIE により $\mathbf{E}[b] = \beta$ を得た. また, Ω 以下のようにかけ, 対称行列である.

$$Var(\epsilon|\mathbf{X}) = \sigma^2 \underbrace{\begin{pmatrix} \omega_1 & \omega_{1,2} & \cdots & \omega_{1,n} \\ \omega_{2,1} & \omega_2 & & \\ & & \ddots & \\ & & & \omega_n \end{pmatrix}}_{\Omega} \quad with \quad \omega_{i,j} = \omega_{j,i} = \frac{Cov(\epsilon_i, \epsilon_j|\mathbf{X})}{\sigma^2}, \omega_i = \frac{Var(\epsilon_i|\mathbf{X})}{\sigma^2}$$

i.i.d. データでは, (A2) を満たすため, Ω は対角行列となる. また, 均一分散の場合は対角成分が全て 1 となる. では問題の分散を考えてみよう. $\mathbf{Var}[\mathbf{b}|\mathbf{X}] = \mathbf{E}[(\mathbf{b} - \beta)(\mathbf{b} - \beta)^T|\mathbf{X}] =$

$\mathbf{E}[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\epsilon\epsilon^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}|\mathbf{X}] = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{E}[\epsilon\epsilon^T|\mathbf{X}]\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\Omega\mathbf{X})(\mathbf{X}^T\mathbf{X})^{-1}$. 誤差項の正規性を仮定するならば, $\mathbf{b}|\mathbf{X} \sim \mathcal{N}(\beta, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\Omega\mathbf{X})(\mathbf{X}^T\mathbf{X})^{-1})$. もはや今までの $s^2(\mathbf{X}^T\mathbf{X})^{-1}$ に基づく議論は無意味である. さらに, 古典的仮定に基づいたガウスマルコフの定理は成り立たず, BULE でももはやない. 再度推定量を BULE にすることを目指すアプローチが一般最小二乗法 (Generalized Least Squares) であり, 再び正確に標準誤差を求めることを目指すアプローチが頑健標準誤差 (Robust standard errors) である.

9.1 一般最小二乗法

一般化回帰モデルの問題は, 不均一分散 (A5) に違反しているために発生する. (A5) が再び満たされるようにデータを変換しよう. Ω が既知とする. すると, $\Omega^{-1} = P^TP$ となる正則行列 P が存在することが示せる. 次に, 回帰方程式に P を掛け, $P\mathbf{Y} = P\mathbf{X}\beta + P\epsilon$ を得る. これを $\mathbf{Y}_\star = \mathbf{X}_\star\beta + \epsilon_\star$ に書き換える. $\epsilon_\star$ の条件付き分散は次のように表される. $\mathbf{Var}[\epsilon_\star|\mathbf{X}_\star] = \mathbf{E}[\epsilon_\star\epsilon_\star^T|\mathbf{X}_\star] = \mathbf{E}[P\epsilon\epsilon^TP^T|\mathbf{X}_\star] = P\mathbf{E}[\epsilon\epsilon^T|\mathbf{X}_\star]P^T = \sigma^2P\Omega P^T = \sigma^2\mathbf{I}$. (A5) が変換後のモデルに再び適用される. 変換後のモデルに従って計算した以下が一般最小二乗推定量である.

$$\mathbf{b}_\star = (\mathbf{X}_\star^T\mathbf{X}_\star)^{-1}\mathbf{X}_\star^T\mathbf{Y}_\star = (\mathbf{X}^T\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}^T\Omega^{-1}\mathbf{Y}$$

(A5) を満たすため, GLSE の $\mathbf{b}_\star$ は再度 BULE となった. 条件付分布は, $\mathbf{b}_\star|\mathbf{X}_\star \sim \mathcal{N}(\beta, \sigma^2(\mathbf{X}_\star^T\mathbf{X}_\star)^{-1})$. ここで, $\mathbf{Var}[\mathbf{b}_\star|\mathbf{X}_\star] = \sigma^2(\mathbf{X}_\star^T\mathbf{X}_\star)^{-1} = \sigma^2(\mathbf{X}^T\Omega^{-1}\mathbf{X})^{-1}$. GLSE は Ω^{-1} で重み付けした, OLSE のような推定量である. しかし, Ω は一般的には**観測不能**. 不均一分散だが系列相関のない特別な場合を考えよう. この結果得られるのが加重最小二乗 (Weighted Least Squares) 推定量である. このとき,

$$\Omega = diag(\omega_1, \omega_2, \dots, \omega_n), \quad \Omega^{-1} = P^TP = diag(\frac{1}{\omega_1}, \frac{1}{\omega_2}, \dots, \frac{1}{\omega_n}), \quad P = diag(\frac{1}{\sqrt{\omega_1}}, \frac{1}{\sqrt{\omega_2}}, \dots, \frac{1}{\sqrt{\omega_n}})$$

$$\because \quad \mathbf{y}_\star = P \cdot \mathbf{y} = \begin{pmatrix} \frac{y_1}{\sqrt{\omega_1}} \\ \frac{y_2}{\sqrt{\omega_2}} \\ \vdots \\ \frac{y_n}{\sqrt{\omega_n}} \end{pmatrix}, \quad \mathbf{X}_\star = P\mathbf{X} = \begin{pmatrix} \frac{\mathbf{x}_1^T}{\sqrt{\omega_1}} \\ \frac{\mathbf{x}_2^T}{\sqrt{\omega_2}} \\ \vdots \\ \frac{\mathbf{x}_n^T}{\sqrt{\omega_n}} \end{pmatrix}, \quad \mathbf{b}_\star = (\sum_{i=1}^n \frac{1}{\omega_i}\mathbf{x}_i\mathbf{x}_i^T)^{-1}(\sum_{i=1}^n \frac{1}{\omega_i}\mathbf{x}_i y_i^T)$$

WLSE を適用すると, 観測値への加重が大きくなるほど, 観測値の分散は小さくなる.

Ex. 時給に関する性差 時給を女性ダミーに回帰して $H_i = \beta_1 + \beta_2F_i + \epsilon_i$ を考える. $Var[\epsilon_i|F_i = 0] = E[(H_i - \beta_1)^2|F_i = 0] = Var[H_i|F_i = 0], Var[\epsilon_i|F_i = 1] = E[(H_i - \beta_1 - \beta_2)^2|F_i = 1] = Var[H_i|F_i = 1]$. ここで, 不均一分散 $Var[\epsilon_i|F_i = 1] \neq Var[\epsilon_i|F_i = 0]$ が認められたとしよう. この場合の適切な推定方法は? **解答** 系列相関を考えなくて良いとして, WLS を利用する. ω_i は $\frac{Var(\epsilon_i|F_i=0)}{\sigma^2}$ か $\frac{Var(\epsilon_i|F_i=1)}{\sigma^2}$ どちらかをとる. ここで, 観測不能な分散を推定量に置き換える必要があり, スカラー σ^2 はすべての重みに共通するため比例関係に影響を与えず取り除く事が出来る. 結果, $\frac{H_i}{\sqrt{\hat{Var}(\epsilon_i|F_i)}}$ を $\frac{1}{\sqrt{\hat{Var}(\epsilon_i|F_i)}}, \frac{F_i}{\sqrt{\hat{Var}(\epsilon_i|F_i)}}$ に回帰することで WLSE$b_\star$ を得る.

9.2 頑健標準誤差

標準誤差に問題があることが分かっているので, これを不均一分散のもとでも使える, すなわち頑健なものに置き換えることは自然な解決策の一つだろう. 簡単のため, ここでも系列相関のない不均一分散を考える. また, 大標本理論に基づく議論であるため, **3.2** での議論と同様 (予定) にデータは i.i.d. かつ (A4a) $\mathbf{E}[\mathbf{x}_i\mathbf{x}_i^T] = Q$ は positive definite, と仮定する. これは正則行列であることの十分条件であり, 正則行列であることは逆行列の存在を保証している. ここで, $\sqrt{n}(\mathbf{b}-\beta) = (\sum_{i=1}^n \frac{1}{n}\mathbf{x}_i\mathbf{x}_i^T)^{-1}\sqrt{n}(\sum_{i=1}^n \frac{1}{n}\mathbf{x}_i\epsilon_i) \xrightarrow{d} \mathcal{N}(\mathbf{0}, Q^{-1}\Sigma Q^{-1})$ だった (?). 均一分散の下では, $\Sigma = \sigma^2\mathbf{E}[\mathbf{x}_i\mathbf{x}_i^T] = \sigma^2 Q$(?). それゆえ, $\mathrm{AVar}[\mathbf{b}] = \frac{\sigma^2}{n}Q^{-1}$, $\widehat{\mathrm{AVar}}[\mathbf{b}] = s^2(\mathbf{X}^T\mathbf{X})^{-1}$. 一方不均一分散の下では, $\Sigma = \mathbf{E}[\epsilon_i^2\mathbf{x}_i\mathbf{x}_i^T]$ であり, $\mathrm{AVar}[\mathbf{b}] = \frac{1}{n}Q^{-1}\Sigma Q^{-1}$. 以前のように Q を $\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T$ で推定し, 同様に $\frac{1}{n-K}\sum_{i=1}^n e_i^2\mathbf{x}_i\mathbf{x}_i^T \xrightarrow{p} \Sigma$ の (左辺) を一致推定量とする, ここで e_i は残差. $\widehat{\mathrm{AVar}}[\mathbf{b}] = \frac{1}{n}(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T)^{-1}(\frac{1}{n-K}\sum_{i=1}^n e_i^2\mathbf{x}_i\mathbf{x}_i^T)(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T)^{-1}$. この頑健推定量は White (1980) によって提案されたもので, 不均一分散の状況に一般的に適応可能. 頑健標準誤差（White standard errors）とホワイトの不均一分散検定は以下の通り.

（**WhiteStandardErrors**） $se(b_k) = \sqrt{\widehat{\mathrm{AVar}}[\mathbf{b}]_{k,k}}$

（**White′sTestforHeteroskedasticity**） $H_0 : \sigma_i^2 = \sigma^2 \forall i \quad vs. \quad H_1 : not \quad H_0$

ホワイト検定の手順: 二乗残差 e_i^2 を X^TX の全ての上三角部分の成分に回帰する. この時検定統計量 nR^2 は漸近的に自由度 $J-1$ の χ^2 に従う. ここで J は定数を含む回帰変数の数.

頑健な標準誤差と検定統計量を計算できる場合, 通常の標準誤差を気にする必要はないのだろうか? 答えは標本サイズに左右される. 頑健検定統計量と信頼区間は, 漸近的な正当性しか持たない. そのため, 標本サイズが小さい場合, 頑健統計量は適切に動作するとは限らず, むしろ通常の統計量よりもバイアスが大きいこともある. ただ, 標本サイズが大きい場合には頑健統計量のみの報告で十分とする見方もある. 最後に, $K=2$ のケースでの公式を求めてみよう. $\sqrt{n}(b-\beta) \xrightarrow{d} \mathcal{N}(0, Q^{-1}\Sigma Q^{-1}) = \mathcal{N}(0, \frac{Var((X_{i,2}-\mu_{X_2})\epsilon_i)}{(Var(X_{i,2}))^2})$, $Var((X_{i,2}-\mu_{X_2})\epsilon_i) = \mathbb{E}[\epsilon_i^2(X_{i,2}-\mu_{X_2})^2]$. 均一分散の時は, $Var((X_{i,2}-\mu_{X_2})\epsilon_i) = \sigma^2 Var(X_2)$. 各部分の一致推定量を利用すれば, 不均一分散の時は $\widehat{\mathrm{AVar}}[b_2] = \frac{1}{n}\frac{\frac{1}{n-2}\sum(X_{i,2}-\bar{X}_2)^2\cdot e_i^2}{(\frac{1}{n}\sum(X_{i,2}-\bar{X}_2)^2)^2} = \frac{1}{n}\frac{\frac{1}{n-2}\sum(X_{i,2}-\bar{X}_2)^2\cdot e_i^2}{(s_{X_2}^2)^2}$, 均一分散の時は $\widehat{\mathrm{AVar}}[b_2] = \frac{1}{n}\frac{s^2\cdot s_{X_2}^2}{(s_{X_2}^2)^2} = \frac{1}{n}\frac{s^2}{s_{X_2}^2}$ である. これはもとの OLS で得られるものと同一.

10 操作変数法

説明変数が誤差項と相関する内生性の存在のもとでは (A1) が崩れ, 正確な因果推論を行えなくなる. 内生性を発生する要因としては, 先述の通り, 欠落変数, 測定誤差, 同時性などがあった. では,

一般に内生性が OLS 推定に引き起こす影響とは? また, コントロール変数, 代理変数のデータ等が入手不可能な場合の解決策とは? ここで我々が導入するのが**操作変数法（Instrumental Variable Method, IV Method）**である. ここでは価格弾力性の推定を例に考えて, その後に一般的な操作変数の定義を行うこととする.

Ex. 需要の価格弾力性と部分均衡　喫煙による病気や死亡を減らすため, タバコに課税したい. タバコの消費量を 20 パーセント削減するには, タバコの販売価格をどのくらい引き上げる必要があるだろうか? これを知るためにタバコの需要の弾力性, つまり**需要方程式**の係数 β_1 を推定する:

$$\underbrace{ln(Q_i)}_{q_i} = \beta_0 + \beta_1 \underbrace{ln(P_i)}_{p_i} + \epsilon_i$$

今, 観測値 $(Q_i, P_i), i = 1, \dots, n$ が得られたとする. OLS 推定は一致性を持つだろうか? 結論としては否, その理由を以下で追っていこう. 部分均衡理論に基づく構造方程式は以下のとおり.

$$\textbf{Demand}: \quad q_i^D = \beta_0 + \underbrace{\beta_1}_{theoretically<0} p_i + \underbrace{\epsilon_i}_{DemandSlide}$$

$$\textbf{Supply}: \quad q_i^S = \alpha_0 + \underbrace{\alpha_1}_{theoretically>0} p_i + \underbrace{\nu_i}_{SupplySlide}$$

各関数 (構造方程式) の誤差項は独立 $Cov(\epsilon_i, \nu_i) = 0$ と仮定する. 均衡価格は市場一掃条件を満たす価格水準として定義されるため, $q_i^S = q_i^D = q_i$ であり, この均衡水準での数量と価格のデータ (q_i, p_i) が n 個与えられていると考えられる. q_i を p_i に回帰したときの OLSE は $b_1^{OLS} = \frac{\hat{Cov}(p_i, q_i)}{\hat{Var}(p_i)} \xrightarrow{p} \beta_1 + \frac{Cov(p_i, \epsilon_i)}{Var(p_i)}$. 均衡点では $\beta_0 + \beta_1 p_i + \epsilon_i = \alpha_0 + \alpha_1 p_i + \nu_i$. 変形して, $p_i = \frac{\alpha_0 - \beta_0 + \nu_i}{\beta_1 - \alpha_1} - \frac{1}{\beta_1 - \alpha_1}\epsilon_i$ with $\beta_1 - \alpha_1 < 0$. 今までの関係式を利用して, $Cov(p_i, \epsilon_i) = -\frac{1}{\beta_1 - \alpha_1} Var(\epsilon_i) > 0$ を得る. b_1^{OLS} の確率極限が $\beta_1 + \frac{Cov(p_i, \epsilon_i)}{Var(p_i)} > \beta_1$ となるため, OLSE は**もはや一致推定量でない**. ここで一致推定量を得る方法として挙がってくるのが**操作変数**の利用だ. 供給関数を以下のように書き換えよう.

$$\textbf{SupplyModified}: \quad q_i^S = \alpha_0 + \alpha_1 p_i + \underbrace{\alpha_2 z_i + u_i}_{\nu_i}$$

ここで, z_i は観測可能で $Cov(z_i, \epsilon_i) = 0$. つまり**供給には影響を与えるが需要には影響を与えない変数**である. $Cov(z_i, u_i) = 0, Cov(\epsilon_i, u_i) = 0$. 均衡では $p_i = \frac{\alpha_0 - \beta_0}{\beta_1 - \alpha_1} + \frac{\alpha_2}{\beta_1 - \alpha_1} z_i + \frac{u_i - \epsilon_i}{\beta_1 - \alpha_1}$ から, $Cov(z_i, p_i) = \frac{\alpha_2}{\beta_1 - \alpha_1} Var(z_i) \neq 0$. 今までの関係式を利用して, $Cov(q_i, z_i) = Cov(\beta_0 + \beta_1 p_i + \epsilon_i, z_i) = \beta_1 Cov(p_i, z_i)$ より, 変形して $\beta_1 = \frac{Cov(q_i, z_i)}{Cov(p_i, z_i)}$ を得る. これを標本対応した $b_1^{IV} = \frac{s_{q,z}}{s_{p,z}}$ が **IV 推定量**である. この推定量は明らかに**一致性を持つ**. 均衡水準は需要, 供給関数両者の位置によって決定するもので, そのスライドを意味する各関数の誤差項の処理が十分でないことに起因する推定のズレが問題である. そのため, **片方の関数の位置を固定し, もう片方の振る舞いを観測データ（均衡水準）から特定できるようにしたい**訳だが, 操作変数法はそれを実現するための手法なのだ.