

末石計量 補足資料

前川 大空 *

2025 年 6 月 12 日

目次

1	線形回帰と OLS	2
1.1	単回帰モデル	2
1.2	重回帰モデル	7
1.3	線形射影	9
2	操作変数法	10
2.1	説明変数と操作変数がひとつの場合の操作変数法	12
2.2	一般の場合の操作変数法	13
2.3	操作変数の見つけ方	14
3	プログラム評価	15
3.1	反実仮想プログラムと実験データ	15
3.2	無視可能性	15
3.3	自然実験と操作変数	16
3.4	LATE	16
3.5	回帰不連続デザイン	17
4	行列表記と漸近理論	17
4.1	準備	17
4.2	OLS 推定量	17
4.3	複合仮説の検定	22
4.4	漸近効率性と GLS 推定量	27
4.5	TSLS 推定量	31
付録 A	計算上のテクニック	33
付録 B	重要な定理等	34

* 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

1 線形回帰と OLS

1.1 単回帰モデル

■p.3 『平均独立は独立性を示唆しない』 まず, 確率変数の独立性の仮定を確認する.

Def: 確率変数の独立性

確率変数 X, Y が $P(XY) = P(X)P(Y)$ ならば, X, Y は独立であるという.

平均独立の仮定より以下を得る:

$$\begin{aligned}\mathbb{E}[u_i] &\stackrel{\text{LIE}}{=} \mathbb{E}[\mathbb{E}[u_i|X_i]] \stackrel{\text{A2}}{=} \mathbb{E}[0] = 0 \\ \therefore \mathbb{E}[u_i] &= \mathbb{E}[u_i|X_i]\end{aligned}$$

また, 以下の関係式も成立する:

$$\begin{aligned}\implies \text{Cov}[u_i X_i] &= \mathbb{E}[u_i X_i] - \mathbb{E}[u_i]\mathbb{E}[X_i] \stackrel{\text{LIE}}{=} \mathbb{E}[X_i \mathbb{E}[u_i|X_i]] - \mathbb{E}[u_i]\mathbb{E}[X_i] \stackrel{\text{A2}}{=} 0 \\ \mathbb{E}[u_i X_i] &= \mathbb{E}[u_i]\mathbb{E}[X_i]\end{aligned}$$

これは独立性の必要条件に過ぎず, 誤差項と説明変数の独立性を意味しない. 仮定は平均が X に依存しないことを意味するのみで, 不均一分散等の可能性は排除しない.

■p.3 単回帰モデルの仮定 1 単回帰モデルの仮定と誤差項の関係について考えたい.

単回帰モデルの仮定 1:

$(X_i, Y_i) \forall i \in \{1, \dots, n\}$ は i.i.d.

u_i の i.i.d. 性も確認する仮定として考えてよいのか? 特に独立性の定義と運用があやふやなので今一度確認したい. $(X_i, Y_i) \forall i \in \{1, \dots, n\}$ は独立. すなわち, 以下が全ての $i \neq j$ で成立していることを指す:

$$\begin{aligned}P((Y_i, X_i) \in A, (Y_j, X_j) \in B) &= P((Y_i, X_i) \in A)P((Y_j, X_j) \in B) \quad \forall A, B \subset \Omega \\ P_{X_i, Y_i, X_j, Y_j}(x_i, y_i, x_j, y_j) &= P_{X_i, Y_i}(x_i, y_i)P_{X_j, Y_j}(x_j, y_j) \quad \forall x_i, y_i \in \Omega\end{aligned}$$

基本的に下の記法を用いて独立性は議論する. 以下では独立を $(X_i, Y_i) \perp (X_j, Y_j)$ とも表記する.

■単回帰モデルにおける独立性 上記で挙げた疑問点に答えを与える. $(X_i, Y_i) \perp (X_j, Y_j)$ から, $(X_i) \perp (X_j)$, $(Y_i) \perp (Y_j)$, $(X_i) \perp (Y_j)$ 等は直ちに言えそうだが, 方針確認のため証明する.

Proof. 定義より以下が成立:

$$P_{X_i, Y_i, X_j, Y_j}(x_i, y_i, x_j, y_j) = P_{X_i, Y_i}(x_i, y_i)P_{X_j, Y_j}(x_j, y_j)$$

上の式と条件付期待値の定義より,

$$\begin{aligned}P_{X_i, Y_i, X_j, Y_j}(x_i, y_i, x_j, y_j) &= P_{X_i, Y_i}(x_i, y_i)P_{X_j, Y_j|X_i, Y_i}(x_j, y_j|x_i, y_i) \\ P_{X_j, Y_j|X_i, Y_i}(x_j, y_j|x_i, y_i) &= P_{X_j, Y_j}(x_j, y_j)\end{aligned}$$

条件付同時 pdf が無条件同時 pdf と一致する。この式自体も独立性の定義の一つ。まずは $(X_i) \perp (X_j)$, 即ち

$$P_{X_j|X_i}(x_j|x_i) \stackrel{\text{def}}{=} \frac{P_{X_i, X_j}(x_i, x_j)}{P_{X_i}(x_i)} = \frac{P_{X_i}(x_i)P_{X_j}(x_j)}{P_{X_i}(x_i)} = P_{X_j}(x_j)$$

を証明したい。ここで,

$$\begin{aligned} P_{X_i, X_j}(x_i, x_j) &= \iint_{\mathbb{R}^2} P_{X_i, Y_i, X_j, Y_j}(x_i, y_i, x_j, y_j) dy_i dy_j \\ &\stackrel{\text{def}}{=} \iint_{\mathbb{R}^2} P_{X_i, Y_i}(x_i, y_i) P_{X_j, Y_j}(x_j, y_j) dy_i dy_j \\ &= \int_{\mathbb{R}} P_{X_i, Y_i}(x_i, y_i) dy_i \int_{\mathbb{R}} P_{X_j, Y_j}(x_j, y_j) dy_j = P_{X_i}(x_i) P_{X_j}(x_j) \end{aligned}$$

よって証明したい等式が成立。他二つも同様に証明が可能。□

次に, 上で示した $(X_i) \perp (X_j)$, $(Y_i) \perp (Y_j)$, $(X_i) \perp (Y_j)$ を用いて, 誤差項の独立性 $(u_i) \perp (u_j)$ を証明したい。

Proof. 単回帰モデルとその仮定:

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \text{ where } (X_i, Y_i) \stackrel{\text{i.i.d.}}{\sim} \forall i \in \{1, \dots, n\}$$

より, $Y_i - \beta_0 - \beta_1 X_i = u_i(X_i, Y_i) = u_i$. つまり, 示したいのは,

$$(u_i) \perp (u_j), \text{ where } u_i = u_i(X_i, Y_i) = Y_i - \beta_0 - \beta_1 X_i$$

$$P_{u_i, u_j}(x_i, y_i, x_j, y_j) = P_{u_i}(x_i, y_i) P_{u_j}(x_j, y_j)$$

$u_i = Y_i - bX_i$ と $u_j = Y_j - bX_j$ との独立性を示すために, 各々の確率密度がどう計算されるか, それらが積の形に分解できるかを考える。 $u_i = z$ の確率に注目する。定義から $u_i = z \implies Y_i = bX_i + z$. $X_i = x$ のとき, $Y_i = bx + z$ となる確率は:

$$P_{u_i}(z) = \int_{\mathbb{R}} P_{X, Y}(x, bx + z) dx$$

という形で書ける。同様に $u_j = w$ となる確率密度は:

$$P_{u_j}(w) = \int_{\mathbb{R}} P_{X, Y}(x', bx' + w) dx'$$

ここで u_i と u_j の同時密度を考えると:

$$\begin{aligned} P_{u_i, u_j}(z, w) &= \iint_{\mathbb{R}^2} P_{X_i, Y_i, X_j, Y_j}(x, bx + z, x', bx' + w) dx dx' \\ &\stackrel{\text{i.i.d.}}{=} \iint_{\mathbb{R}^2} P_{X_i, Y_i}(x, bx + z) P_{X_j, Y_j}(x', bx' + w) dx dx' \\ &= \int_{\mathbb{R}} P_{X, Y}(x, bx + z) dx \int_{\mathbb{R}} P_{X, Y}(x', bx' + w) dx' \stackrel{\text{def}}{=} P_{u_i}(z) P_{u_j}(w) \end{aligned}$$

よって目的の関係式を得た。□

よって, $(u_i) \perp (u_j)$ が確認できた。以上の議論より, 誤差項同士の独立性は仮定 1 のみで保証可能である。しかし, 同一分布性は先述の通り保証されず, 以下のように結論づけられる。

単回帰モデルの仮定下での誤差項の性質:

単回帰モデルの仮定の下で, 誤差項は独立だが i.i.d. ではない。

■p.4 識別と A2 観測可能な変数 (Y_i, X_i) の同時分布が分かれば、というのがミソ。識別の定義: 『観測されるデータの分布が既知の時, θ の値が一意に定まるならば, θ は識別されるという』に沿った記述であることを確認すること。より詳しく言えば, (X_i, Y_i) の同時分布が判明することで, $\mathbb{E}[Y_i|X_i]$ が X_i の関数として決定される。ここで (1.2) 式は X_i の一変数関数として考えられるため, β が一意に定まる (=識別される)。

■p.6 『漸近分散』の語の用法 自分の Note での用法と違うことに注意。状況によって柔軟に判断しよう。例えば, p.6 や p.9 で出てくる漸近分散は正確に言えば $\sqrt{n}(\hat{\beta}_1 - \beta_1)$ のものである。対して標準誤差という語は $\hat{\beta}_1$ そのものに用いられていることを注意しよう。

■p.6 $\sqrt{n}(\hat{\beta}_1 - \beta_1)$ の厳密分布 中級計量で与えられた説明は, 古典的仮定, 特に均一分散と正規性 (条件付きで良い) の下での議論。 $\sqrt{n}(\hat{\beta}_1 - \beta_1)$ を標準化したものは, 分散が観測可能な場合は標準正規分布に, 不可能な場合には自由度 $n-2$ の t 分布に従っていた。この仮定を外した場合の議論は p.9 で少し展開されている。重要なのは漸近理論が適応可能で, この漸近分布はどの場合でも正規分布であることだ。

■p.8 『i.i.d. 仮定下で誤差項の無条件分散はすべての i について必ず等しくなる』

Proof. 以下の単回帰モデルと仮定のもとで議論を行う:

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \text{ where } (X_i, Y_i) \text{ は RVs, } \overset{\text{i.i.d.}}{\sim} \forall i \in \{1, \dots, n\}$$

$\text{Var}(u_i)$ が i に関し一定であると示す。 $\text{Var}(Y_i) = \sigma_y^2, \text{Var}(X_i) = \sigma_x^2$ とすると: $\sigma_y^2 = \text{Var}(\beta_1 X_i + u_i) = \beta_1^2 \sigma_x^2 + \text{Var}(u_i) + 2\beta_1 \text{Cov}(X_i, u_i), (\text{const.}) = \text{Var}(u_i) + 2\beta_1 \text{Cov}(X_i, u_i) \stackrel{\text{A2}}{=} \text{Var}(u_i)$. \square

不均一分散の定義: —

単回帰モデルの仮定の下では無条件分散の不均一分散は実現せず, p.7 の形で定義を行う必要がある。

■p.9 不均一分散下での $\hat{\beta}_1$ の標準誤差導出 ここでは OLSE の分散を求めて, その標本対応から導いた標準誤差が古典的仮定のもとでの標準誤差とは異なることを示す。単回帰モデルとその仮定:

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \text{ where } (X_i, Y_i) \overset{\text{i.i.d.}}{\sim} \forall i \in \{1, \dots, n\}$$

OLS 推定量は

$$(\hat{\beta}_0, \hat{\beta}_1) = (\bar{Y} - \hat{\beta}_1 \bar{X}, \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}) = (\beta_0 + (\beta_1 - \hat{\beta}_1) \bar{X} + \bar{u}, \beta_1 + \frac{\sum_{i=1}^n (X_i - \bar{X}) u_i}{\sum_{i=1}^n (X_i - \bar{X})^2})$$

まず不偏性を示そう。

Proof.

$$\begin{aligned} \mathbb{E}[\hat{\beta}_1] &= \beta_1 + \mathbb{E}\left[\frac{\sum_{i=1}^n (X_i - \bar{X}) u_i}{\sum_{i=1}^n (X_i - \bar{X})^2}\right] \\ &\stackrel{\text{LIE}}{=} \beta_1 + \mathbb{E}\left[\mathbb{E}\left[\frac{\sum_{i=1}^n (X_i - \bar{X}) u_i}{\sum_{i=1}^n (X_i - \bar{X})^2} \middle| \mathbf{X}\right]\right] \\ &= \beta_1 + \mathbb{E}\left[\frac{\sum_{i=1}^n (X_i - \bar{X}) \mathbb{E}[u_i | \mathbf{X}]}{\sum_{i=1}^n (X_i - \bar{X})^2}\right] \\ &\stackrel{\text{i.i.d.}}{=} \beta_1 + \mathbb{E}\left[\frac{\sum_{i=1}^n (X_i - \bar{X}) \mathbb{E}[u_i | X_i]}{\sum_{i=1}^n (X_i - \bar{X})^2}\right] \stackrel{\text{A2}}{=} \beta_1 \end{aligned}$$

$\hat{\beta}_0$ も LIE により不偏性を確認できる. □

$Var(\hat{\beta}_1) = \mathbb{E}[(\hat{\beta}_1 - \beta_1)^2] = \mathbb{E}[(\frac{\sum_{i=1}^n (X_i - \bar{X})u_i}{\sum_{i=1}^n (X_i - \bar{X})^2})^2] \stackrel{\text{LIE}}{=} \mathbb{E}[\frac{\mathbb{E}[(\sum_{i=1}^n (X_i - \bar{X})u_i)^2 | \mathbf{X}]}{[\sum_{i=1}^n (X_i - \bar{X})^2]^2}]$. 期待値内の分母を考えよう.

$$(\text{denominator}) = \mathbb{E}[(\sum_{i=1}^n (X_i - \bar{X})u_i)^2 | \mathbf{X}] = \mathbb{E}[S^2 | \mathbf{X}] \quad \text{where } S = \sum_{i=1}^n (X_i - \bar{X})u_i$$

$$\mathbb{E}[S^2 | \mathbf{X}] = \sum_{i=1}^n (X_i - \bar{X})^2 \mathbb{E}[u_i^2 | \mathbf{X}] + 2 \sum_{i < j} (X_i - \bar{X})(X_j - \bar{X}) \mathbb{E}[u_i u_j | \mathbf{X}].$$

先述の議論より, A1 より **誤差項同士が独立**, つまり **相関がない** (\because 独立性の必要条件) ので:

$$\mathbb{E}[u_i u_j | \mathbf{X}] = \mathbb{E}[u_i | \mathbf{X}] \mathbb{E}[u_j | \mathbf{X}] = 0, \quad \therefore \mathbb{E}[S^2 | \mathbf{X}] = \sum_{i=1}^n (X_i - \bar{X})^2 \mathbb{E}[u_i^2 | \mathbf{X}] = \sigma_i^2 \quad \forall i \neq j$$

$$Var(\hat{\beta}_1) = \mathbb{E}[\frac{\sum_{i=1}^n (X_i - \bar{X})^2 \mathbb{E}[u_i^2 | \mathbf{X}]}{[\sum_{i=1}^n (X_i - \bar{X})^2]^2}]$$

$$Var(\hat{\beta}_1 | \mathbf{X}) \stackrel{\text{LIE}}{=} \frac{\sum_{i=1}^n (X_i - \bar{X})^2 \mathbb{E}[u_i^2 | \mathbf{X}]}{[\sum_{i=1}^n (X_i - \bar{X})^2]^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 \sigma_i^2}{[\sum_{i=1}^n (X_i - \bar{X})^2]^2}$$

更に均一分散の仮定を置けば:

$$Var(u_i | X_i) = \mathbb{E}[u_i^2 | \mathbf{X}] = \sigma^2 \quad \forall i, \quad \therefore \mathbb{E}[S^2 | \mathbf{X}] = \sigma^2 \sum_{i=1}^n (X_i - \bar{X})^2$$

$$Var(\hat{\beta}_1) = \sigma^2 \mathbb{E}[\frac{1}{\sum_{i=1}^n (X_i - \bar{X})^2}], \quad Var(\hat{\beta}_1 | \mathbf{X}) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

古典的仮定下での OLSE の **条件付分散** が得られ, これは本文の記述と一致している. だが実際は $Var(u_i | X_i) = \sigma(X_i)$ と X_i に依存することがしばしばで, この場合古典的仮定下での標準誤差は推定には使えない.

■ 『単回帰モデルの仮定』 下での漸近分散 条件付分散の確率収束先, 漸近分散の振る舞いを見よう.

Proof. 漸近分布という語は標準化を暗黙に要求していたことに注意して:

$$\begin{aligned} V = Var(\sqrt{n}(\hat{\beta}_1 - \beta_1) | \mathbf{X}) &= \frac{n \sum_{i=1}^n [(X_i - \bar{X})^2 \sigma_i^2]}{[\sum_{i=1}^n (X_i - \bar{X})^2]^2} \\ &\stackrel{\text{heterogeneity}}{=} \frac{\frac{1}{n} \sum_{i=1}^n [(X_i - \bar{X})^2 \sigma^2(X_i)]}{[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2]^2} \\ &\xrightarrow{p} \frac{\mathbb{E}[(X_i - \mathbb{E}[X])^2 \sigma^2(X_i)]}{\sigma_X^4} \\ &= \frac{\mathbb{E}[(X_i - \mathbb{E}[X])^2 \mathbb{E}[u_i^2 | \mathbf{X}]]}{\sigma_X^4} \\ &\stackrel{\text{LIE}}{=} \frac{\mathbb{E}[(X_i - \mathbb{E}[X])^2 u_i^2]}{\sigma_X^4} = \frac{Var((X_i - \mathbb{E}[X])u_i)}{Var(X_i)^2} \end{aligned}$$

ここでは連続写像定理, 4 次モーメントまでの有限性を用いた. 教科書の記述と合致する. □

■誤差項に関する期待値計算の性質 上記の導出において用いた誤差項に関する期待値計算の性質の正統性を確認しておこう。まず、不偏性の証明で使った $\mathbb{E}[u_i|\mathbf{X}] \stackrel{\text{i.i.d.}}{=} \mathbb{E}[u_i|X_i]$ を確認する。次に、 $\mathbb{E}[u_i u_j|\mathbf{X}] = \mathbb{E}[u_i|X_i]\mathbb{E}[u_j|X_j]$ と、この結果に必要な仮定を clarify しておこう。

Proof. $\mathbb{E}[u_i | \mathbf{X}] = \mathbb{E}[u_i | X_i]$ を示したい。つまり以下が成り立つことを確かめたい:

$$\begin{aligned}\mathbb{E}[u_i | \mathbf{X}] &= \int u_i f_{u_i|\mathbf{X}}(u_i | \mathbf{X}) du_i \\ &\stackrel{(\dagger)}{=} \int u_i f_{u_i|X_i}(u_i | X_i) du_i = \mathbb{E}[u_i | X_i].\end{aligned}$$

ここで (†) を示したい。単回帰モデルの仮定

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad (X_i, Y_i) \stackrel{\text{i.i.d.}}{\sim}$$

より $u_i = u_i(X_i, Y_i) = Y_i - \beta_0 - \beta_1 X_i$ は X_i と Y_i で決定され、 \mathbf{X}_{-i} は一切関与しない。よって、

$$f_{u_i|\mathbf{X}}(u_i | \mathbf{X}) = f_{u_i|X_i, \mathbf{X}_{-i}}(u_i | X_i, \mathbf{X}_{-i}) = f_{u_i|X_i}(u_i | X_i),$$

が成り立つため、(†) も示され題意が満たされた。□

Proof. 次に、 $\mathbb{E}[u_i u_j | \mathbf{X}] = \mathbb{E}[u_i | X_i] \mathbb{E}[u_j | X_j]$, $i \neq j$ を検証しよう。下を示すのが目標だ:

$$\begin{aligned}\mathbb{E}[u_i u_j | \mathbf{X}] &= \iint u_i u_j f_{u_i, u_j|\mathbf{X}}(u_i, u_j | \mathbf{X}) du_i du_j \\ &\stackrel{(\ddagger)}{=} \iint u_i u_j f_{u_i|X_i}(u_i | X_i) f_{u_j|X_j}(u_j | X_j) du_i du_j \\ &= \left(\int u_i f_{u_i|X_i}(u_i | X_i) du_i \right) \left(\int u_j f_{u_j|X_j}(u_j | X_j) du_j \right) = \mathbb{E}[u_i | X_i] \mathbb{E}[u_j | X_j].\end{aligned}$$

ここで (‡) を示したい。先に示した誤差項の独立性と、 (X_i, Y_i) の i.i.d. 性から

$$f_{u_i, u_j|\mathbf{X}} = f_{u_i|X_i} f_{u_j|X_j}.$$

によって (‡) が成立し、したがって題意は満たされた。□

■連続写像定理の正統性検証 次に、漸近分布の導出で用いた連続写像定理 (CMT) の正統性を検証しよう。

Thm: 連続写像定理 (Continuous Mapping Theorem)

確率収束 $X_n \xrightarrow{P} X$ を満たす確率変数列 $\{X_n\}$ と、関数 $g: \mathbb{R}^k \rightarrow \mathbb{R}$ が点 x の近傍で連続であるとき、

$$g(X_n) \xrightarrow{P} g(X) \quad (\text{ただし } X_n \xrightarrow{P} X).$$

では CMT を確認したところで、目的の式変形の正統性を確認しよう。なお、分母における分散の確率収束については以前検証したため、ここでは確認せずに確率収束を認める。

Proof. 標準化を暗黙に含むことに注意して、以下の分子・分母それぞれの確率収束を示す

$$V = \text{Var}(\sqrt{n}(\hat{\beta}_1 - \beta_1) | \mathbf{X}) = \frac{\frac{1}{n} \sum_{i=1}^n [(X_i - \bar{X})^2 \sigma^2(X_i)]}{[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2]^2}.$$

$$\begin{aligned}
(\text{denominator}) &= A_n^2 = \left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^2 \xrightarrow{P} (\text{Var}(X_i))^2 = \sigma_X^4 \quad (\because \text{LLN, CMT}) \\
(\text{numerator}) &= B_n = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \sigma^2(X_i) = \frac{1}{n} \sum_{i=1}^n f(X_i, \bar{X}), \\
&\text{where } f: \mathbb{R}^2 \rightarrow \mathbb{R}, \quad f(x, m) = (x - m)^2 \sigma^2(x).
\end{aligned}$$

ここで $\sigma(x)$ の連続性を仮定すれば f は連続関数である。各 i について、 $(X_i, \bar{X}) \xrightarrow{P} (X_i, \mathbb{E}[X])$ である（第一成分は定常に同じ分布、第二成分は LLN による）。CMT より

$$f(X_i, \bar{X}) \xrightarrow{P} f(X_i, \mathbb{E}[X]) = (X_i - \mathbb{E}[X])^2 \sigma^2(X_i).$$

さらに、 X_i の i.i.d. 性と 4 次モーメントまでの有限性の下で LNN を再適用すると、

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n f(X_i, \bar{X}) &\xrightarrow{P} \mathbb{E}[f(X_i, \mathbb{E}[X])] = \mathbb{E}[(X_i - \mathbb{E}[X])^2 \sigma^2(X_i)]. \\
B_n &\xrightarrow{P} \mathbb{E}[(X_i - \mathbb{E}[X])^2 \sigma^2(X_i)]
\end{aligned}$$

これらを組み合わせ、LLN, CMT, LIE を用いて目的の確率収束先を得た。□

1.2 重回帰モデル

■p.15 完全な多重共線性 ある説明変数が他の説明変数に線形従属である状態のことを指す。『R による実証分析』勉強会の 11, 12 章の資料でも散々議論したのでそちらも参照してほしい。今回は OLS 推定量の導出に注目した議論を、注 7 に絡めてしてみたい。（線形）回帰モデルの OLSE は以下の問題を解き与えられる：

$$\begin{aligned}
\text{重回帰モデル: } \min_{\mathbf{b}} \sum_i [Y_i - \mathbb{E}[Y_i | \mathbf{X}_i]]^2 &= \min_{\mathbf{b}} \sum_i [Y_i - \mathbf{X}_i \mathbf{b}]^2 \\
\text{単回帰モデル: } \min_{b_0, b_1} \sum_i [Y_i - \mathbb{E}[Y_i | X_i]]^2 &= \min_{b_0, b_1} \sum_i [Y_i - b_0 - b_1 X_i]^2
\end{aligned}$$

説明変数行列 \mathbf{X} の第一列（それゆえ行ベクトル \mathbf{X}_i の第一成分も）の要素は全て 1 であることに注意せよ。単回帰モデルは説明変数が定数項と X の 2 つのみの重回帰モデルと言える。この最小化問題で重要なのは、解が一意に定まるかである。注 7 の通り、変数 X が一定になる場合、つまり：

$$b_0 + b_1 X_i = b_0 \cdot 1 + b_1 \cdot c = \hat{y} \text{ (constant)}$$

のとき、我々の目的は先述の損失関数を最小化することにあったが、一定の \hat{y} に対して我々が選択できるのは (b_0, b_1) の二つである。これは不定方程式であり、 \mathbf{b} の選び方は無数に存在するため識別が不可能である。さらにこの場合、説明変数ベクトル X は定数項ベクトル $\mathbf{1}$ と完全な多重共線性を持つことが分かる。つまり：

$$X = \begin{pmatrix} c \\ \vdots \\ c \end{pmatrix} = c \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = c \cdot \mathbf{1}$$

以上より、我々がよく知る OLSE を得るためには、以下のように条件付けをするといいいことが分かった：

単回帰モデルの OLS 推定量

説明変数 X が一定でないとき (= 定数項との間に完全な多重共線性がないとき)，

$$(\hat{\beta}_0, \hat{\beta}_1) = (\bar{Y} - \hat{\beta}_1 \bar{X}, \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}).$$

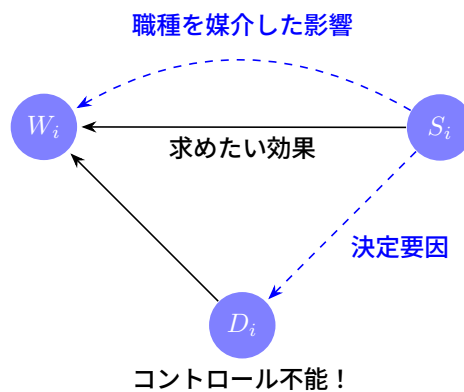
同様の議論が重回帰モデルの議論でも成り立つ。^{*1} 末石計量では行列表記が 4 章以降に回されるので、ここでは結果だけ示しておこう。導出は Econometrics Note で行っており、記法もこれに従う。^{*2}

重回帰モデルの OLS 推定量

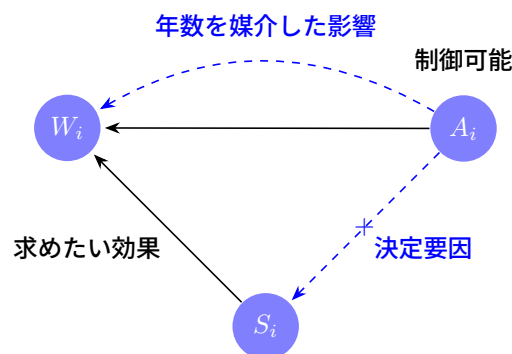
説明変数間に完全な多重共線性がないとき，

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}.$$

■p.16 不適切なコントロール変数 本文にあげられた例を図示してみよう。



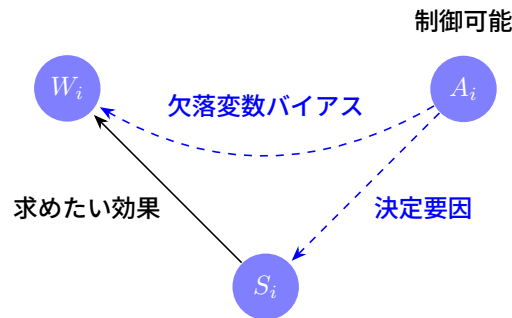
能力がコントロール変数として使えるのが何故か、職種との差異にも一考の余地がある。この例では能力がコントロール可能であるため、能力から教育年数への影響をうまくつぶすことができる。



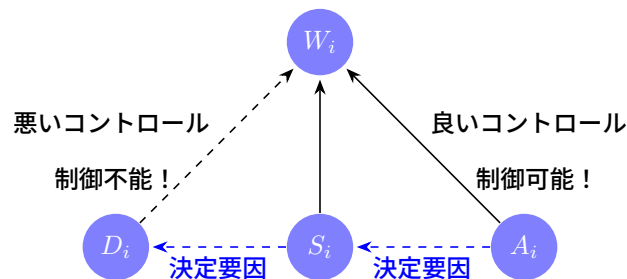
^{*1} 繰り返しになるが、Dummy Variable Trap の議論はまさに多重共線性について言及したものである。

^{*2} OLSE に使用される行列 $\mathbf{X}^T \mathbf{X}$ に逆行列が存在するのは列フルランク行列のみであるため、 $\text{rank}(\mathbf{X}) = \text{rank}(\mathbf{X}^T \mathbf{X})$ で、列ベクトルが線形独立でない (完全な多重共線性が存在する) \Leftrightarrow 正則ではない \Leftrightarrow 逆行列が存在しない。このように OLS 推定量の導出に用いられる公式は完全な多重共線性の下では機能しないが、そもそもモデル仮定の時点でこのようなケースは弾かれるのである。

欠落変数バイアスは以下のような図示が可能だ。



関係を整理するとこの通り。興味のある説明変数の決定要因になるものはコントロール変数として利用できるが、興味のある説明変数が決定要因になるものは利用できない (bad control)。



■p.17 不完全な多重共線性 Econometrics Note では冗長なコントロール変数の問題と絡めて、説明変数の相関関係が大きいときの問題を指摘したが、本文の記述の通りこれは一般的にも成り立つ。分散が大きくなってしまうものの、明確な対処法は存在しないため気にしすぎるものではない。

1.3 線形射影

■p.19 不完全な多重共線性 (未解決) 最適化問題の建付けとして、線形射影は期待値の最小化である一方で、回帰モデルではわざわざ期待値の (おそらく) 標本対応を使って議論を始めているのは何故だろうか。

$$\text{線形射影モデル: } \min_{\mathbf{b}} \mathbb{E}[Y_i - \mathbf{X}_i \mathbf{b}]^2, \mathbb{E}[u_i | \mathbf{X}_i] \neq 0$$

$$\text{重回帰モデル: } \min_{\mathbf{b}} \sum_i [Y_i - \mathbb{E}[Y_i | \mathbf{X}_i]]^2 = \min_{\mathbf{b}} \sum_i [Y_i - \mathbf{X}_i \mathbf{b}]^2, \mathbb{E}[u_i | \mathbf{X}_i] = 0$$

一致推定できる、との結論に帰着するため？ 4.3 章の記述を見る限り、線形射影モデルは母集団について考えているようだ。 **というか線形射影モデルが『予測』を全部担ってるのかな？**

■p.20 標本対応 標本対応は一致性により正当化される。Econometrics Note の 4 ページを確認すること。

■p.20 線形モデルの有用性 線形射影は Y_i のみならず、 $\mathbb{E}[Y_i | \mathbf{X}_i]$ の線形最良近似である。一つのデータのみならず、分布についての情報を与えてくれる点で線形モデルは有用だと言える。

2 操作変数法

■p.21 欠落変数バイアスと内生性の違い 流れを Clarify しておこう.

実際のモデル: $Y_i = \beta_0 + \beta_1 S_i + \beta_2 A_i + u_i$, where A_i : unobservable, β_1 : interested, $\mathbb{E}[u_i | S_i, A_i] = 0$

単回帰モデル: $Y_i = \gamma_0 + \gamma_1 S_i + \epsilon_i$, where $\gamma_1 \neq \beta_1$: NOT interested (biased), $\mathbb{E}[\epsilon_i | S_i] = 0$

推定を目指した立式: $Y_i = \beta_0 + \beta_1 S_i + \hat{u}_i$, where $\hat{u}_i = \beta_2 A_i + u_i$, BUT, $\mathbb{E}[\hat{u}_i | S_i] \neq 0$: (A2) failed!

2 つ目の式が欠落変数バイアスで, 3 つ目の式が内生性の例である. 内生性は回帰モデルの仮定に反しており, 最早回帰モデルとしての議論の上で教育年数の賃金への効果が語れなくなっている. ^{*3} これを無視して OLS の "公式" に従って計算してみると, 結果は結局二つ目の推定と同じになってしまう. ^{*4}

■p.22 内生変数の定義と由来 Econometric Note の操作変数法の項に記載した部分均衡分析の例を再利用してみよう. 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

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

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 ε_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.

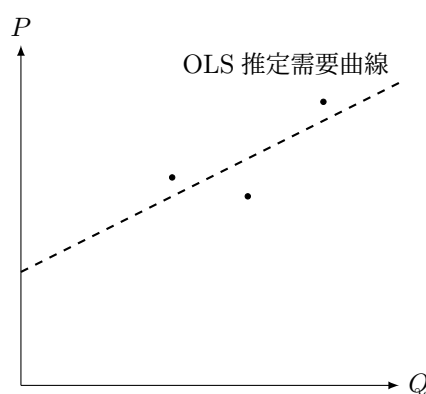


図 1 誤った推定による右上がりの需要曲線

^{*3} 回帰の建付けを思い出してみると, $Y_i = \mathbb{E}[Y_i | \mathbf{X}_i] + (Y_i - \mathbb{E}[Y_i | \mathbf{X}_i]) := \mathbb{E}[Y_i | \mathbf{X}_i] + \epsilon_i$. LIE から (A2) は自然と成り立つ. (A2) に違反していることはそもそもこの構成を否定しているため, 回帰とは言えないはずだ.

^{*4} Econometric Note にも同じ流れに沿って記載した筈だが, 3 つ目の議論が拙いので参考にしない方がいい....

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。

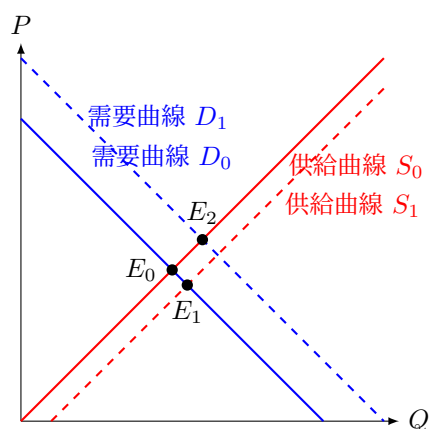


図2 外生ショックによるシフト

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 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。

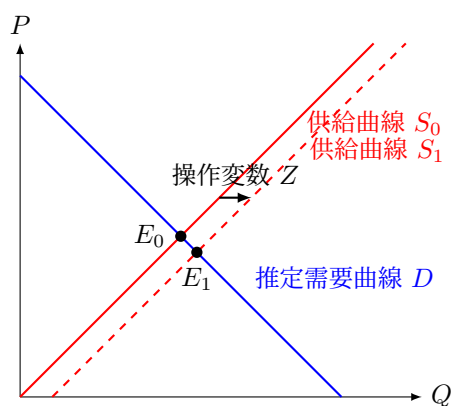


図3 操作変数 Z による供給曲線のシフト

最後に、本例における内生・外生を整理する。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 ε_d 、供給ショック ε_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。

2.1 説明変数と操作変数がひとつの場合の操作変数法

■p.22 内生性と相関 $Cov(X_i, \varepsilon_i) \neq 0$ の仮定は、平均独立の仮定の否定の十分条件であることが、補足資料 p.1 の対偶を取ってやればわかる。また、線形射影モデルは平均独立の仮定の代わりに $E[X_i u_i] = 0, E[u_i] = 0$ を仮定して構成されていたことを思い出せば、この構成が線形射影モデルですらないことも分かる。つまり：

内生性の位置づけ

線形回帰モデル: $E[Y_i|X_i] = \beta_0 + \beta_1 X_i$ と実際の生成過程が線形で、 $E[u_i|X_i] = 0$

線形射影モデル: 生成過程に関わらない線形最良近似、 $E[u_i|X_i] \neq 0$ だが $E[X_i u_i] = 0, E[u_i] = 0$

内生性のある構造型: $Cov(X_i, \varepsilon_i) \neq 0$ で上記のどちらの仮定も満たさないが、モデルを記述する

■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。

■p.24 除外制約 部分均衡分析の例でいえば、操作変数 Z は Q^d に直接影響を与えない、つまり、需要ショック ε_d に含まれない必要が、需要曲線を固定するためにあると言える。

■p.24 二段階最小二乗法 他の推定方法である 2SLS の流れを整理して、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、 $\hat{P}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$ を得る（誘導形）。
2. 第二段階：得られた予測値 \hat{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 β_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 $\nu_i = P_i - \hat{P}_i$ 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。

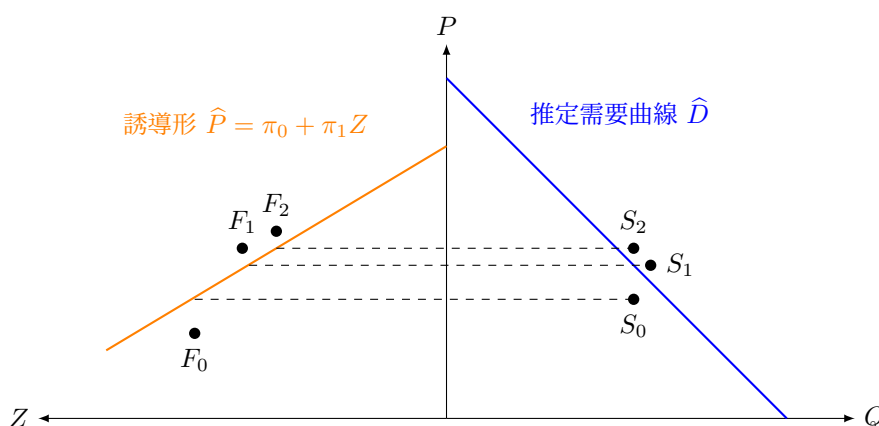


図 4 2SLS の流れ

■p.24 誘導形と構造型 構造型が主体の意思決定を表現したモデルであるのに対し、誘導形は内生変数を外生変数で表した形のことを指す。Z は先述の通り外生変数であるため、第一段階は誘導型であることが分かる。なお、ここでの定義から分かるように、構造型か否かは内生性とは独立したものである。しかし、誘導形は線形射影モデルの範疇であり、識別の問題は果たされる。

■p.25 IV と 2SLS の差異 IV は丁度識別でのみ使えるが、2SLS は過剰識別の場合も使える点で強力である。

■p.26 Wald 推定量 操作変数がダミー変数の場合は Wald 推定量で簡単に効果を推測可能。

$$\text{IV がダミーの場合の推定対象: } \beta_1 = \frac{\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0]}{\mathbb{E}[X_i | Z_i = 1] - \mathbb{E}[X_i | Z_i = 0]}$$

後述するが、Local Average Treatment Effect (LATE) が、潜在効果モデルでの対応する結果である。

2.2 一般の場合の操作変数法

■p.28 内生変数がある場合の関連性 考えるのは以下の構造モデル:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_{1,i} + \beta_3 W_{2,i} + \cdots + \beta_{1+r} W_{r,i} + e_i$$

で、 \mathbf{W} はコントロール変数、つまりそれ自体に興味はないが制御の為に回帰に組み込まれる外生変数である。ここで X_i は内生変数で、 e_i と相関をもつ。これを解決するために操作変数を用いた誘導形が導入される。(2.11) 式と (2.13) 式を比較することにより、連立方程式の形に識別の問題を帰着させることが可能で:

$$(2.11) \text{ 式の係数} = \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_r \\ \gamma_{r+1} \\ \vdots \\ \gamma_{r+m} \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 \pi_0 \\ \beta_2 + \beta_1 \pi_1 \\ \vdots \\ \beta_{r+1} + \beta_1 \pi_r \\ \beta_1 \pi_{r+1} \\ \vdots \\ \beta_1 \pi_{r+m} \end{pmatrix} = (2.13) \text{ 式の係数},$$

ここで、 $(\beta_0, \dots, \beta_{r+1})$ は未知数である。 β を一意に定める条件を詳細に考えると、 $(\pi_{r+1}, \dots, \pi_{r+m})$ が一つでも 0 でなければ、 $\beta_1 = \gamma_{r+k} / \pi_{r+k}$ に確定する。ここで連立方程式の上 r 行に立ち戻ると、 β_1 が確定したことから β_0 が定まり、同様に $\beta_2, \dots, \beta_{r+1}$ も確定する。これを整理して関連性の条件は得られる。

関連性

内生変数 X_i の誘導形:

$$X_i = \pi_0 + \pi_1 W_{1,i} + \cdots + \pi_r W_{r,i} + \pi_{r+1} Z_{r+1,i} + \cdots + \pi_{r+m} W_{m,i} + \nu_i$$

での操作変数 \mathbf{Z} にかかる係数 $(\pi_{r+1}, \dots, \pi_{r+m})$ は零ベクトルではない。

■次数条件, 階数条件 構造型の識別の必要条件は次数条件 (丁度 or 過剰識別), 必要十分条件は階数条件。

2.3 操作変数の見つけ方

■AK(1991) のパズル 能力と教育年数に正の相関があるならば、能力の項を落とし、誤差に入れ込んでしまった OLS では係数を過剰推定するだろうとの予想を欠落変数バイアスの定式化から確認することが出来る。『パズル』とは、予想とは逆に、OLS では過少推定されていることを意味する。IV 法の肝を明示的に示そう：

操作変数法での推定

IV 法では、相関とそれに応じたバイアスの方向を予測して、実データからその関係性が実際に成り立つのかを検証することが重要である。

また、この要因として教育効果の異質性が挙げられている。3 章での LATE に繋がる議論だろう。

■操作変数と内生変数 関連性は必ずしも内生変数と操作変数の間の因果関係を意味しない。これは IV 法にて説明変数が IV による誘導型で表現されたことから確認できる。従って、**相関するだけの要因も IV に利用できるだろう**。例えば、『R による実証分析』11 章の出席率が成績に与える影響の実証分析については、定期価格 (通学時間が決定要素の一つ、だが出席率にはあまり因果効果がなさそう) を IV として用いることもできるだろう。通学時間↑ → 定期価格↑, 出席率↓で定期=出席に負の相関関係 (×因果) があるが、定期価格は学業成績には影響を与えない点で外生性も満たすだろう。(例として問題ない?)

■p.33 弱操作変数法 教科書に記載のある例での識別の条件を確認する：

構造型： $Y_i = \beta_1 X_i + e_i$, where $\mathbb{E}[e_i] = 0$

誘導型： $X_i = \pi_1 Z_i + \nu_i$, where $\mathbb{E}[\nu_i Z_i] = 0$, $\mathbb{E}[\nu_i] = 0$,

誘導型 $Y_i = \gamma_1 X_i + \mu_i$, where $\mathbb{E}[\mu_i X_i] = 0$, $\mathbb{E}[\mu_i] = 0$ を考える。構造型に X_i の誘導型を代入すると：

誘導型： $Y_i = \beta_1 \pi_1 Z_i + \beta_1 \nu_i + e_i$, where $\mathbb{E}[(\beta_1 \nu_i + e_i) Z_i] = 0$, $\mathbb{E}[\beta_1 \nu_i + e_i] = 0$,

を Z_i の外生性 ($Cov(Z_i, e_i) = 0$) から得る。ここで、誘導型と構造型の係数を比較すれば：

$$\gamma_1 = \beta_1 \pi_1$$

誘導形は識別されるので、この構造型係数 β_1 の識別には、 $\pi_1 \neq 0$ があれば保証される。よって、記載された事項を確認できた。次に、TSLS 推定量を計算する。切片無し回帰より：

$$\begin{aligned} & \min_{\pi_1^0} \sum_i [X_i - \pi_1^0 Z_i]^2, \\ \text{FOC: } & \sum_i Z_i [X_i - \pi_1^0 Z_i] = 0, \quad \sum_i Z_i X_i = \sum_i Z_i [\hat{\pi}_1 Z_i] = \sum_i Z_i \hat{X}_i \\ & \hat{\beta}_1^{TSLS} = \frac{\sum_i \hat{X}_i Y_i}{\sum_i \hat{X}_i^2} \\ & = \frac{\sum_i Z_i Y_i}{\sum_i Z_i \hat{X}_i} \quad (\because \hat{X}_i = \hat{\pi}_1 Z_i) \\ & = \frac{\sum_i Z_i Y_i}{\sum_i Z_i X_i} \quad (\because \sum_i Z_i X_i = \sum_i Z_i \hat{X}_i) \end{aligned}$$

この TSLS 推定量は教科書に記載されているものと合致している。なお、定数無しの単回帰は $\beta_0 = 0$ を暗に仮定していることになり、これが崩れた場合には OLS 推定にバイアスが発生する。

■p.33 注7について 不正確なのは π_1 の扱いだろう。識別の十分条件である $\pi_1 \neq 0$ に違反する点で怪しい (注8を見る限り実際に識別されない)。更に弱操作変数を考えるなら、右側極限 $\lim_{\pi_1 \downarrow 0}$ を考えるべきだ。

3 プログラム評価

3.1 反実仮想プログラムと実験データ

■p.37 回帰分析, 反実仮想 回帰分析と、プログラム評価のうち反実仮想フレームワークの差異を整理する。

回帰分析: $Y_i = \beta_0 + \beta_1 S_i + \beta_2 A_i + u_i$, where β_1 : interested

反実仮想フレームワーク: $Y_i = D_i Y_{1,i} + (1 - D_i) Y_{0,i}$, where $Y_{1,i} - Y_{0,i}$: interested

この二つについて重要な差異は、何が観測不可能なのかにある。

2つの観測不可能性

フレームワークによって何が観測不可能なのか、それに伴い生起する性質は異なる:

回帰分析: 構造型に入れ込まれた変数の内いくつか、内生性の原因となりうる

反実仮想: $(Y_{0,i}, Y_{1,i})$ の一方 (『因果推論の根本問題』), 異質性を許容し, 各 ATE の端緒

■p.38 識別と観測可能性 識別の概念は、観測されるデータに関する概念! 今回は:

観測可能: $f_{Y_1|D}(y|1), f_{Y_0|D}(y|0)$

観測不可能: $f_{Y_1|D}(y|0), f_{Y_0|D}(y|1), f_{Y_1}(y), f_{Y_0}(y)$

観測不可能な分布が対象に含まれる場合は識別されようがないわけだ。そこで、反実仮想フレームワークにおいては、推定対象を全て観測可能なもので表示することが目標で、そのための仮定が今後3章では提示される。

■p.40 ATE と OLS トリートメント効果が個人ごとに異なる, $\beta_{1,i} = Y_{1,i} - Y_{0,i}$ で特徴づけられる回帰モデルでは、推定対象である $\mathbb{E}[\beta_{1,i}]$ の OLS は ATE の一致推定量となる。ここで注意すべきなのは、 $\mathbb{E}[u_i | D_i] = 0$ が建付けから成立しており、回帰モデルとして成立している事だろう。『Rによる実証分析』8章の補足を参照した記述だが、推定対象は個人ごと? 若しくは平均?

3.2 無視可能性

■p.42 ATE の導出 LIE で導出できることを確かめよう。

$$\text{ATE} = \mathbb{E}[Y_{1,i} - Y_{0,i}] = \mathbb{E}[\mathbb{E}[Y_{1,i} - Y_{0,i} | X_i]]$$

これを X_i の実現値について書き下せば教科書の記述を得る。次に、 $\text{ATE}(x_j)$ も確認してみよう。

$$\begin{aligned} \text{ATE}(x_j) &= \mathbb{E}[Y_{1,i} - Y_{0,i} | X_i = x_j] \\ &= \sum_{D_i \in \{0,1\}} P(D_i | X_i = x_j) \mathbb{E}[Y_{1,i} - Y_{0,i} | D_i, X_i = x_j] \end{aligned}$$

ここから教科書の記述を得る。

■p.42 無視可能性と観測可能性 今回、コントロールの対象は観測可能なものを想定していることに注意を払う必要がある。ここで、コントロールしたいものの観測が出来ないなら、操作変数と LATE が議論に挙がる。

■p.43 コントロール変数の条件 処置を受ける人と受けない人がともに存在しなければならないのは、完全な多重共線性を排除するための仮定だろう。

■p.43 重回帰モデルと無視可能性 $\mathbb{E}[u_i | D_i, X_i] = \mathbb{E}[u_i | X_i]$ が無視可能性 $(Y_{0,i}, Y_{1,i}) \perp\!\!\!\perp D_i | X_i$ から成り立つことを確かめたい。以下のように確認できる：

$$\begin{aligned} \mathbb{E}[u_i | D_i, X_i] &\stackrel{(\text{LIE})}{=} \mathbb{E}[\mathbb{E}[u_i | Y_{0,i}, Y_{1,i}, D_i, X_i] | D_i, X_i] \\ &\stackrel{(\text{無視可能性})}{=} \mathbb{E}[\mathbb{E}[u_i | Y_{0,i}, Y_{1,i}, X_i] | D_i, X_i] = \mathbb{E}[u_i | X_i]. \end{aligned}$$

次に、重回帰モデルの仮定を満たすことを確認したい。

$$\begin{aligned} u_i &= \mathbb{E}[u_i | X_i] + \eta_i \\ \mathbb{E}[Y_{0,i}] &= \mathbb{E}[\beta_0 + u_i] = \beta_0 + \mathbb{E}[u_i] = \beta_0, \because \mathbb{E}[u_i] = 0, \text{ おそらく必要な追加的仮定} \\ \mathbb{E}[u_i | X_i] &= \mathbb{E}[Y_{0,i} | X_i] - \beta_0 = (\alpha_0 + \alpha_1 X_i) - \beta_0, \\ Y_i &= \beta_0 + \beta_1 D_i + u_i = \beta_0 + \beta_1 D_i + [(\alpha_0 + \alpha_1 X_i) - \beta_0] + \eta_i \\ &= (\alpha_0 + \beta_0) + \beta_1 D_i + \alpha_1 X_i - [\beta_0] + \eta_i \\ &= (\alpha_0 + \beta_0 - \mathbb{E}[Y_{0,i}]) + \beta_1 D_i + \alpha_1 X_i + \eta_i + [-\beta_0 + \mathbb{E}[Y_{0,i}]] \\ &= (\alpha_0 + \beta_0 - \mathbb{E}[Y_{0,i}]) + \beta_1 D_i + \alpha_1 X_i + \eta_i \\ &= \gamma_0 + \beta_1 D_i + \alpha_1 X_i + \eta_i, \quad \gamma_0 \equiv \alpha_0 + \beta_0 - \mathbb{E}[Y_{0,i}]. \end{aligned}$$

これなんで $(\alpha_0 + \beta_0 - \mathbb{E}[Y_{0,i}])$ として長ったらしい表記をしてるんだ? $= \alpha_0$ だけど?

$$\begin{aligned} \mathbb{E}[\eta_i | D_i, X_i] &= \mathbb{E}[(u_i - \mathbb{E}[u_i | X_i]) | D_i, X_i] \\ &= \mathbb{E}[(u_i - \mathbb{E}[u_i | D_i, X_i]) | D_i, X_i] \\ &= \mathbb{E}[u_i | D_i, X_i] - \mathbb{E}[\mathbb{E}[u_i | D_i, X_i] | D_i, X_i] \\ &= \mathbb{E}[u_i | D_i, X_i] - \mathbb{E}[u_i | D_i, X_i] = 0 \end{aligned}$$

より、重回帰モデルの仮定も確認でき、一致推定を保証することが出来た。

3.3 自然実験と操作変数

■p.45 操作変数と異質性 この節では処置効果を同質なものとして仮定しており、2SLS での処置効果が可能であることは記載しているが、仮想現実フレームワークからのアプローチは明示されていない。

3.4 LATE

■p.47 Wald 推定量と LATE 2 章で登場した Wald 推定量の公式と 3 章の LATE についての差異を考える。二つの推定対象には、公式には明示的に表れていないが、重要な差異として D_i の異質性がある。この差異は仮

定を比べることで確認することが可能である。また、基づくモデルはそれぞれ順に回帰と反実仮想 FW.

$$\begin{aligned} \text{IV がダミーの場合の推定対象: } \mathbb{E}[\beta_{1,i}] &= \frac{\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0]}{\mathbb{E}[X_i | Z_i = 1] - \mathbb{E}[X_i | Z_i = 0]} \\ \text{LATE} = \mathbb{E}[Y_{1,i} - Y_{0,i} | D_{1,i} > D_{0,i}] &= \frac{\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0]}{\mathbb{E}[D_i | Z_i = 1] - \mathbb{E}[D_i | Z_i = 0]} \end{aligned}$$

推定対象は平均? 若しくは個人のもの?

■p.48 操作変数と異質性 重要な差異として D_i の異質性を挙げたが、この重要性は、教科書に記載されているように、操作変数の取り方によって条件 $D_{1,i} > D_{0,i}$ に適合する集団 (**Complier**) が変化することにある。

$$\text{LATE} = \mathbb{E}[Y_{1,i} - Y_{0,i} | \mathbf{D}_{1,i} > \mathbf{D}_{0,i}]$$

LATE は Complier に対する平均的な処置効果を表しているため、異質性によって、操作変数の取り方によって推定対象は変動する事を意識する必要がある。

3.5 回帰不連続デザイン

■p.49 RDD と重回帰モデル 交差項のない重回帰モデルを考えているが、切片のみのシフトとは限らないはずだ。もう少し一般的な説明は『R による実証分析』10 章に記載されている。例えば、結果の解釈は閾値の近傍でのみ有効であることに注意する必要がある。

■p.50 sharp RDD と fuzzy RDD 推定方法の差異に注意しよう。sharp RDD では完全にコントロール可能であることから、3.2 章の議論を適応して OLS で推定が可能である。一方で fuzzy RDD は操作変数としての利用となることから、3.4 章の議論を適応して 2SLS を利用する必要がある。

4 行列表記と漸近理論

4.1 準備

■p.53 delta method 中心極限定理を、標本平均の関数に対しても利用できるようになる、みたいな定理。

4.2 OLS 推定量

■p.55 正則と識別 教科書に沿って重回帰モデル

$$Y_i = \mathbf{X}_i' \beta + u_i$$

を考え、識別の対象である β が如何なるものかを確認しよう。このままでは誤差項を消す操作を行ったところでも β にかかる \mathbf{X}_i' を取り除くことができない。そこで、両辺に \mathbf{X}_i を掛けて、期待値をとると以下のように(逆行列をとりうる) 正方行列の形で書ける:

$$\begin{aligned} \mathbb{E}[\mathbf{X}_i Y_i] &= \mathbb{E}[\mathbf{X}_i \mathbf{X}_i'] \beta + \mathbb{E}[\mathbf{X}_i u_i] \\ &= \mathbb{E}[\mathbf{X}_i \mathbf{X}_i'] \beta \quad (\because A2) \end{aligned}$$

ここで $\mathbb{E}[\mathbf{X}_i \mathbf{X}_i']^{-1}$ を計算するためには正則性が要求される。 (\mathbf{X}_i, Y_i) の同時分布が観測できるとき、上記の式で登場する期待値は全て一意に定まり、正則の場合は逆行列を両辺に掛けることで解くことが出来るため、明

らかに識別される。問題は、正則でなく、逆行列が存在しないときである。上記の式は未知数 β についての連立方程式となる。ここでは、行列 $A = \mathbb{E}[\mathbf{X}_i \mathbf{X}_i'] \in \mathbb{R}^{(K+1) \times (K+1)}$ と表記し、 $\beta \in \mathbb{R}^{K+1}$ である。

Thm: Rank-Nullity Theorem

任意の行列 $A \in \mathbb{R}^{(K+1) \times (K+1)}$ に対して、次が成り立つ：

$$\underbrace{\text{rank}(A)}_{\substack{\text{次元} \\ (=r)}} + \underbrace{\dim(\ker(A))}_{\substack{\text{核空間の次元} \\ (=s)}} = \underbrace{K+1}_{\substack{\text{列数 (未知数の個数)} \\ = \dim(\beta)}}.$$

ここで

$$\ker(A) = \{v \in \mathbb{R}^{K+1} \mid Av = \mathbf{0}\}, \quad \dim(\ker(A)) = s.$$

$$A\beta = b, \quad b = \mathbb{E}[\mathbf{X}_i Y_i],$$

を満たす $\beta \in \mathbb{R}^{K+1}$ を考える。行列 A のランクを r ，核空間の次元は $\dim(\ker(A)) = s$ とすると。

1. $r = K + 1$: A は正則行列である。よって $\beta = A^{-1}b$ と識別可能。
2. $r < K + 1$: $\dim(\ker(A)) = (K + 1) - r > 0$ 。 $A\beta = b$ の解は無数に存在し、一意でない。

$r = K + 1$ のケースの議論を詳しく見よう：

何らかの「特定解」 β^* を一つ見つけて (Ex. 母集団の真の値 β_0) 代入してみる：

$$A\beta^* = b.$$

次に、行列 A の核空間 $\ker(A)$ の基底を以下のようにとる：

$$\langle v_1, v_2, \dots, v_s \rangle \quad (s = (K + 1) - r, Av_j = \mathbf{0}).$$

行列 $V = [v_1, v_2, \dots, v_s] \in \mathbb{R}^{(K+1) \times s}$ を定義して、 $\alpha = (\alpha_1, \dots, \alpha_s)' \in \mathbb{R}^s$ を任意に動かして、

$$\begin{aligned} \beta &= \beta^* + V\alpha \\ A\beta &= A\beta^* + (AV)\alpha \\ &= A\beta^* + \mathbf{0}\alpha = b \end{aligned}$$

を満たす。逆に、 $A\beta = b$ を満たす任意の β は上記の形に一意に書ける。したがって、解集合は

$$\mathcal{B} = \left\{ \beta^* + V\alpha \mid \alpha \in \mathbb{R}^s \right\},$$

で表される。特に、 $\dim(\mathcal{B}) = s > 0$ であるため、 β は一意には定まらず自由度 s をもつ。

■p.56 β の推定 計算は Econometrics Note で詳細に記載している。

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} = \left(\sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top \right)^{-1} \sum_{i=1}^n \mathbf{X}_i Y_i.$$

■p.57 射影行列 以下で定義される射影行列について、性質を考える：

$$\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$$

以下のように教科書の記述を確認できる:

$$\begin{aligned}\mathbf{P}_X \mathbf{Y} &= \mathbf{X}[(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}] = \mathbf{X} \hat{\beta} = \hat{\mathbf{Y}} \\ (\mathbf{I} - \mathbf{P}_X) \mathbf{Y} &= \mathbf{Y} - \mathbf{P}_X \mathbf{Y} = \mathbf{Y} - \hat{\mathbf{Y}} = \hat{\mathbf{u}}\end{aligned}$$

■p.57 不偏性 単回帰モデルで行った補足的説明をなぞる形で確認できる.

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top (\mathbf{X} \beta + \mathbf{u}) \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{X}) \beta + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{u} \\ &= \beta + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{u} = \beta + \left(\sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top \right)^{-1} \sum_{i=1}^n \mathbf{X}_i u_i\end{aligned}$$

Proof. さて, 変形した式を用いて不偏性を確認しよう.

$$\begin{aligned}\mathbb{E}[\hat{\beta}] &= \beta + \mathbb{E}\left[\left(\sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top\right)^{-1} \sum_{i=1}^n \mathbf{X}_i u_i\right] \\ &\stackrel{\text{LIE}}{=} \beta + \mathbb{E}\left[\mathbb{E}\left[\left(\sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top\right)^{-1} \sum_{i=1}^n \mathbf{X}_i u_i \mid \mathbf{X}\right]\right] \\ &\stackrel{\text{誤差項の独立}}{=} \beta + \mathbb{E}\left[\left(\sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top\right)^{-1} \sum_{i=1}^n \mathbf{X}_i \mathbb{E}[u_i \mid \mathbf{X}_i]\right] \stackrel{\text{A2}}{=} \beta\end{aligned}$$

平均独立の仮定 (A2) が不偏性のために重要であることが分かる. 注 1 にも書いていた通り. \square

■p.57 一致性

$$\hat{\beta} = \beta + \left(\sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top \right)^{-1} \sum_{i=1}^n \mathbf{X}_i u_i = \beta + \left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top \right)^{-1} \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i u_i$$

i.i.d. 性から $\mathbf{X}_i \mathbf{X}_i^\top$, $\mathbf{X}_i u_i$ も i.i.d. (後半は要確認) である. 任意の行列に対し, フロベニウス・ノルム

$$\|\mathbf{A}\| := \sqrt{\text{tr}(\mathbf{A}^\top \mathbf{A})},$$

が定義され, $\text{tr}(\cdot)$ はトレースを表す. また, 積が正方行列となる任意の行列 \mathbf{A}, \mathbf{B} について以下が成り立つ:

$$\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$$

$\mathbf{A} = \mathbf{X}_i \mathbf{X}_i^\top$ の場合を考える:

$$\begin{aligned}\|\mathbf{X}_i \mathbf{X}_i^\top\| &= \sqrt{\text{tr}[(\mathbf{X}_i \mathbf{X}_i^\top)^\top (\mathbf{X}_i \mathbf{X}_i^\top)]} \\ &= \sqrt{\text{tr}(\mathbf{X}_i \mathbf{X}_i^\top \mathbf{X}_i \mathbf{X}_i^\top)} \quad (\because \mathbf{A} = \mathbf{A}^\top) \\ &= \sqrt{\text{tr}[\mathbf{X}_i (\mathbf{X}_i^\top \mathbf{X}_i \mathbf{X}_i^\top)]} \\ &= \sqrt{\text{tr}[(\mathbf{X}_i^\top \mathbf{X}_i \mathbf{X}_i^\top) \mathbf{X}_i]} \quad (\because \text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})) \\ &= \sqrt{\text{tr}[(\mathbf{X}_i^\top \mathbf{X}_i) (\mathbf{X}_i^\top \mathbf{X}_i)]} \quad ((\mathbf{X}_i^\top \mathbf{X}_i \mathbf{X}_i^\top \mathbf{X}_i) : \text{スカラー}) \\ &= \sqrt{(\mathbf{X}_i^\top \mathbf{X}_i)} \sqrt{(\mathbf{X}_i^\top \mathbf{X}_i)} \quad (\mathbf{X}_i^\top \mathbf{X}_i : \text{スカラー}) \\ &= \sqrt{\text{tr}(\mathbf{X}_i^\top \mathbf{X}_i)} \sqrt{\text{tr}(\mathbf{X}_i^\top \mathbf{X}_i)} = \|\mathbf{X}_i\|^2.\end{aligned}$$

上式と, (A3) より以下が示される:

$$\mathbb{E}[\|\mathbf{X}_i \mathbf{X}_i^\top\|] = \mathbb{E}[\|\mathbf{X}_i\|^2] < \infty.$$

次に, 以下のように $A_n \in \mathbb{R}^{(K+1) \times (K+1)}$ を置く:

$$A_n := \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top$$

先に示した独立性と有限性から, 大数の法則を適用できる. すなわち,

$$\frac{1}{n} \sum_{i=1}^n [\mathbf{X}_i \mathbf{X}_i^\top]_j \xrightarrow{p} \mathbb{E}[(\mathbf{X}_i \mathbf{X}_i^\top)_j], \quad \forall \text{列 } j = 1, \dots, (K+1).$$

このことから行列として

$$A_n = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top \xrightarrow{p} A = \mathbb{E}[\mathbf{X}_i \mathbf{X}_i^\top],$$

が成立する. ただし, 行列の確率収束 (要素ごとの確率収束) を意味する. (A4) 後半より $A = \mathbb{E}[\mathbf{X}_i \mathbf{X}_i^\top]$ が正則から, 逆行列 A_n^{-1} が存在する. さらに CMT から

$$A_n^{-1} = \left(\frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} \xrightarrow{p} (\mathbb{E}[\mathbf{X}_i \mathbf{X}_i^\top])^{-1} = A^{-1}. \quad (*)$$

次に, 以下のように $b_n \in \mathbb{R}^{(K+1)}$ を置く:

$$b_n := \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i u_i$$

ここで, 先に独立性は示したため, あとは $\mathbb{E}[u_i^2]$, $\mathbb{E}[\|X_i u_i\|] < \infty$ を示せば, LLM が適用可能だ:

まず, $\mathbb{E}[u_i^2] < \infty$ の確認をしよう:

$$u_i = Y_i - \mathbf{X}_i^\top \beta.$$

スカラーでのノルムの扱いに注意して, Minkowski の不等式を用いると,

$$\begin{aligned} (\mathbb{E}[\|u_i\|^p])^{1/p} &= (\mathbb{E}[\|Y_i + (-\mathbf{X}_i^\top \beta)\|^p])^{1/p} \\ &\leq (\mathbb{E}[\|Y_i\|^p])^{1/p} + (\mathbb{E}[\|-\mathbf{X}_i^\top \beta\|^p])^{1/p} \\ &\leq (\mathbb{E}[\|Y_i\|^p])^{1/p} + (\mathbb{E}[\|\mathbf{X}_i\|^p])^{1/p} \|\beta\|. \end{aligned}$$

最後の不等号では Cauchy-Schwarz の不等式を用いた. ここで, $p = 2$ を代入して:

$$(\mathbb{E}[\|u_i\|^2])^{1/2} \leq (\mathbb{E}[\|Y_i\|^2])^{1/2} + (\mathbb{E}[\|\mathbf{X}_i\|^2])^{1/2} \|\beta\|.$$

(A3) より

$$\mathbb{E}[Y_i^2] < \infty, \quad \mathbb{E}[\|\mathbf{X}_i\|^2] < \infty.$$

よって

$$\sqrt{\mathbb{E}[u_i^2]} \leq \sqrt{\mathbb{E}[Y_i^2]} + \sqrt{\mathbb{E}[\|\mathbf{X}_i\|^2]} \|\beta\| < \infty, \quad \therefore \mathbb{E}[u_i^2] < \infty.$$

次に, $\mathbb{E}[\|\mathbf{X}_i u_i\|] < \infty$ を Cauchy-Schwarz の不等式を用いて示す.

$$\begin{aligned}\mathbb{E}[\|\mathbf{X}_i u_i\|] &= \mathbb{E}[|u_i| \|\mathbf{X}_i\|] \\ &\leq \sqrt{\mathbb{E}[u_i^2]} \sqrt{\mathbb{E}[\|\mathbf{X}_i\|^2]}.\end{aligned}$$

$\mathbb{E}[u_i^2] < \infty$ は上で示した通り. かつ仮定 3 より $\mathbb{E}[\|\mathbf{X}_i\|^2] < \infty$ が成り立つので, 右辺は有限. したがって,

$$\mathbb{E}[\|\mathbf{X}_i u_i\|] < \infty.$$

以上より, 大数の法則を適用できる条件が満たされた. よって, LLN より:

$$b_n = \frac{1}{n} \sum_{i=1}^n (\mathbf{X}_i u_i) \xrightarrow{p} \mathbb{E}[\mathbf{X}_i u_i] = \mathbf{0}, \quad (**)$$

が成り立つ. 最後に利用した $\mathbb{E}[\mathbf{X}_i u_i] = \mathbf{0}$ は (A2) の必要条件であり, (A2) が一致性には必要ない.

以上の結果 ((*) と (**)) を代入すると,

$$\hat{\beta} = \beta + A_n^{-1} b_n.$$

右辺の第 2 項 $A_n^{-1} b_n$ について, 確率収束に関する性質により:

$$A_n^{-1} b_n \xrightarrow{p} A^{-1} \mathbf{0} = \mathbf{0}.$$

したがって, $\hat{\beta} = \beta + A_n^{-1} b_n$ において

$$\hat{\beta} \xrightarrow{p} \beta.$$

これで最小二乗推定量 $\hat{\beta}$ の一致性が示された. 教科書では CMT と確率収束を $\mathbf{g}(A, b)$ で一挙に考えている.

■p.59 漸近正規性 最後に, OLSE $\hat{\beta}$ の漸近正規性を示し, 対応する共分散行列とその推定量を導出する.

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{X}_i u_i \right).$$

ここで $\{\mathbf{X}_i u_i\}$ は i.i.d. で, $\mathbb{E}[\mathbf{X}_i u_i] = \mathbf{0}$. ここで, Cauchy-Schwarz の不等式と (A3) より:

$$\mathbb{E}[\|\mathbf{X}_i u_i\|^2] \leq \sqrt{\mathbb{E}[u_i^4]} \sqrt{\mathbb{E}[\|\mathbf{X}_i\|^4]} < \infty$$

が成り立つため, 中心極限定理より:

$$\begin{aligned}\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{X}_i u_i &\xrightarrow{d} N(\mathbf{0}, \mathbb{E}[\mathbf{X}_i \mathbf{X}_i^\top u_i^2]) \\ \left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top \right)^{-1} &\xrightarrow{p} (\mathbb{E}[\mathbf{X}_i \mathbf{X}_i^\top])^{-1}.\end{aligned}$$

更に, 一致性で用いた (*) と組合わせて Slutsky の定理を適用すると,

$$\begin{aligned}\sqrt{n}(\hat{\beta} - \beta) &= \left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{X}_i u_i \right) \\ &\xrightarrow{d} (\mathbb{E}[\mathbf{X}_i \mathbf{X}_i^\top])^{-1} N(\mathbf{0}, \mathbb{E}[\mathbf{X}_i \mathbf{X}_i^\top u_i^2]) \\ &\stackrel{d}{=} N\left(0, (\mathbb{E}[\mathbf{X}_i \mathbf{X}_i^\top])^{-1} \mathbb{E}[\mathbf{X}_i \mathbf{X}_i^\top u_i^2] (\mathbb{E}[\mathbf{X}_i \mathbf{X}_i^\top])^{-1}\right).\end{aligned}$$

したがって、最終的に以下を得る:

$$\mathbf{V} = \mathbb{E}[\mathbf{X}_i \mathbf{X}_i^\top]^{-1} \mathbb{E}[\mathbf{X}_i \mathbf{X}_i^\top u_i^2] (\mathbb{E}[\mathbf{X}_i \mathbf{X}_i^\top])^{-1}$$

これを、 $\hat{\beta}$ の漸近分散共分散行列と呼ぶ。実際のデータでは標本対応して推定される。すなわち:

$$\hat{\mathbf{V}} = \left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top \hat{u}_i^2 \right) \left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top \right)^{-1}$$

これは、White の頑健標準誤差推定量の多変量版である。各回帰係数の標準誤差は、 $\hat{\mathbf{V}}/n$ の対角成分の平方根によって与えられる。

さらなる仮定として条件付き均一分散, すなわち $\mathbb{E}[u_i^2 | X_i] = \sigma^2$ (constant) が成り立つとすると,

$$\mathbb{E}[X_i X_i' u_i^2] = \mathbb{E}[X_i X_i' \sigma^2] = \sigma^2 \mathbb{E}[X_i X_i'] = \sigma^2 A.$$

したがって、漸近分散共分散行列は

$$\mathbf{V} = A^{-1} (\sigma^2 A) A^{-1} = \sigma^2 A^{-1}.$$

さらに、残差の自由度を補正するため、 σ^2 の代わりに

$$s^2 = \frac{1}{n - (k+1)} \sum_{i=1}^n \hat{u}_i^2$$

を用いれば、対応する分散共分散行列推定量は以下の通り:

$$\hat{\mathbf{V}} = s^2 \left(\frac{1}{n} X' X \right)^{-1}$$

繰り返しになるが、不均一分散そのものに興味がない限りは頑健標準誤差推定量を利用すべきである。

■A3 の活用法 p.4 で記載されている、 $\mathbb{E}[|X|^k] < \infty \implies \mathbb{E}[|X|^s] < \infty$ where $s \leq k$ を暗黙に利用して漸近理論の議論は進んでいることに注意しよう。

4.3 複合仮説の検定

■p.61 『統計検定量は既知の分布に収束』 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。故に『既知の分布に分布収束する』という条件は帰無仮説の下では定義から必ず満たされねばならず、そのために標準化は必要不可欠だ。

■p.61 Wald 統計量の構成 複合仮説 (joint hypothesis) の帰無仮説 $H_0: \mathbf{r}(\beta) = \mathbf{0}$, where $\mathbf{r}: \mathbb{R}^{k+1} \rightarrow \mathbb{R}^q$ を考える。帰無仮説とのずれが大きいきこれを棄却する、故に標準化した $\|\mathbf{r}(\hat{\beta})\|^2$ を考えればよい。

$$W_n = n \cdot \mathbf{r}(\hat{\beta})^\top \cdot (\hat{\mathbf{R}} \hat{\mathbf{V}} \hat{\mathbf{R}}^\top)^{-1} \cdot \mathbf{r}(\hat{\beta})$$

$\hat{\mathbf{V}}$ は $\hat{\beta}$ の漸近分散共分散行列推定量, $\hat{\mathbf{R}} = \partial \mathbf{r}(b) / \partial b^\top |_{b=\hat{\beta}} \in \mathbb{R}^{q \times (k+1)}$. 検定統計量かを漸近分布の計算で確かめる。帰無仮説の下 $\mathbf{r}(\hat{\beta}) \xrightarrow{P} \mathbf{0}$ (\cdot : $\hat{\beta}$ の一致性, ベクトル関数の連続性から CMT) より, デルタ法から:

$$\sqrt{n}(\mathbf{r}(\hat{\beta}) - \mathbf{r}(\beta)) = \sqrt{n}(\mathbf{r}(\hat{\beta}) - \mathbf{0}) = \sqrt{n} \cdot \mathbf{r}(\hat{\beta}) \xrightarrow{d} \mathcal{N}_q(\mathbf{0}, \mathbf{RVR}^\top),$$

$\mathbf{R} = \partial \mathbf{r}(b) / \partial b^\top |_{b=\beta}$. ベクトル関数 $\mathbf{r}(\cdot)$ の微分可能性とそれに伴い保証される連続性が利用されていることに注意しよう。 $\hat{\mathbf{R}} \hat{\mathbf{V}} \hat{\mathbf{R}}^\top$ は \mathbf{RVR}^\top の標本対応より, $\mathbf{r}(\hat{\beta})$ の漸近分散共分散行列の一致推定量である。次に, $\hat{\mathbf{R}} \hat{\mathbf{V}} \hat{\mathbf{R}}^\top$ に平方根が存在することを確認したいが、これには次のことを確認する必要がある。

■共分散行列は非負値定符号行列 n 次元確率ベクトル $\mathbf{X}^\top = (X_1, \dots, X_n)$ を考える．期待値は:

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

共分散行列は $\mathbf{V} = \mathbf{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top]$ で与えられる． $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$ から，共分散行列は対称行列であり， $\mathbf{V} = \mathbf{E}[\mathbf{X}\mathbf{X}^\top] - \boldsymbol{\mu}\boldsymbol{\mu}^\top$ である．次に，任意の定数ベクトル $\mathbf{a} = (a_1, \dots, a_n)^\top \in \mathbb{R}^n$ を取ると， $z = \mathbf{a}^\top \mathbf{X}$ は確率変数のスカラーで， z の平均は以下で与えられる:

$$\mathbb{E}(z) = a_1 \mathbb{E}(X_1) + \dots + a_n \mathbb{E}(X_n) = \mathbf{a}^\top \mathbf{E}(\mathbf{X}) = \mathbf{a}^\top \boldsymbol{\mu}$$

分散は,

$$\begin{aligned} \text{Var}(z) &= \text{Var}(\mathbf{a}^\top \mathbf{X}) = \mathbb{E}((\mathbf{a}^\top \mathbf{X} - \mathbf{a}^\top \boldsymbol{\mu})(\mathbf{a}^\top \mathbf{X} - \mathbf{a}^\top \boldsymbol{\mu})^\top) \\ &= \mathbb{E}[\mathbf{a}^\top (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top \mathbf{a}] \\ &= \mathbf{a}^\top \mathbf{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top] \mathbf{a} = \mathbf{a}^\top \mathbf{V} \mathbf{a} \quad (\mathbf{V} \text{ の 2 次形式}) \end{aligned}$$

$\text{Var}(z) \geq 0$ ゆえ，共分散行列 \mathbf{V} は非負値定符号行列である．この構造は分散の標本対応でも保持される．サイズ m のサンプル $\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(m)}\}$ を考える．ここで各 $\mathbf{x}^{(k)} \in \mathbb{R}^n$ は母集団から独立に抽出された観測ベクトルの実現値であるとする．標本平均ベクトル $\bar{\mathbf{x}}$ を以下のように定義する:

$$\bar{\mathbf{x}} = \frac{1}{m} \sum_{k=1}^m \mathbf{x}^{(k)}.$$

標本共分散行列 $\widehat{\mathbf{V}}$ を以下のように定義する:

$$\begin{aligned} \widehat{\mathbf{V}} &= \frac{1}{m} \sum_{k=1}^m (\mathbf{x}^{(k)} - \bar{\mathbf{x}})(\mathbf{x}^{(k)} - \bar{\mathbf{x}})^\top \in \mathbb{R}^{n \times n}, \\ \widehat{\mathbf{V}}_{i,j} &= \frac{1}{m} \sum_{k=1}^m (x_i^{(k)} - \bar{x}_i)(x_j^{(k)} - \bar{x}_j), \quad i, j = 1, 2, \dots, n. \end{aligned}$$

さて，標本共分散行列が非負値定符号であることの証明を考えよう． $z = \mathbf{a}^\top \mathbf{x}^{(k)}$ の標本平均は

$$\frac{1}{m} \sum_{k=1}^m \mathbf{a}^\top \mathbf{x}^{(k)} = \mathbf{a}^\top \left(\frac{1}{m} \sum_{k=1}^m \mathbf{x}^{(k)} \right) = \mathbf{a}^\top \bar{\mathbf{x}}.$$

故に各標本の偏差は $\mathbf{a}^\top \mathbf{x}^{(k)} - \mathbf{a}^\top \bar{\mathbf{x}} = \mathbf{a}^\top (\mathbf{x}^{(k)} - \bar{\mathbf{x}})$ と表記でき， z の標本分散は次のように表される:

$$\begin{aligned} \widehat{\text{Var}}(z) &= \frac{1}{m} \sum_{k=1}^m (\mathbf{a}^\top (\mathbf{x}^{(k)} - \bar{\mathbf{x}}))^2 \\ &= \frac{1}{m} \sum_{k=1}^m (\mathbf{a}^\top (\mathbf{x}^{(k)} - \bar{\mathbf{x}}))(\mathbf{a}^\top (\mathbf{x}^{(k)} - \bar{\mathbf{x}}))^\top \\ &= \mathbf{a}^\top \left[\frac{1}{m} \sum_{k=1}^m (\mathbf{x}^{(k)} - \bar{\mathbf{x}})(\mathbf{x}^{(k)} - \bar{\mathbf{x}})^\top \right] \mathbf{a} = \mathbf{a}^\top \widehat{\mathbf{V}} \mathbf{a} \quad (\widehat{\mathbf{V}} \text{ の 2 次形式}). \end{aligned}$$

$\widehat{\text{Var}}(z) \geq 0$ で，任意のベクトル \mathbf{a} で成り立つため，標本共分散行列 $\widehat{\mathbf{V}}$ も非負値定符号行列である．

■直交対角化と正定値行列の検証 長々と検証したが、実際はもう少し簡単に $\hat{\mathbf{V}} \geq 0$ は検証が可能だ。標本共分散行列 $\hat{\mathbf{V}}$ は定義から対称行列である。対称行列は直交行列による対角化が可能で、その対角成分は固有値に対応する。共分散行列の場合、固有値は各方向の分散を表しており、分散は常に非負であるので、対角成分はすべて非負。よって対称かつ固有値が非負であることから、 $\hat{\mathbf{V}}$ は正規直交行列 \mathbf{Q} を用いて:

$$\hat{\mathbf{V}} = \mathbf{Q} \text{diag}(\lambda_1, \dots, \lambda_n) \mathbf{Q}^\top, \quad \lambda_i \geq 0$$

の形に対角化でき、これは非負値定符号行列であることと同値である。^{*5}

■p.62 $\hat{\mathbf{R}}\hat{\mathbf{V}}\hat{\mathbf{R}}^\top$ に平方根が存在する 本題に戻ろう。標本分散は非負値定符号行列であることが分かった。

引用: 末石本の行列平方根についての記述

$A \geq 0$ のとき、 $A = BB^\top$ を満たすような行列 B が存在し、 B は A の平方根と呼ばれる。 A の平方根は $A^{1/2}$ と表す。 $A^{1/2}$ は必ずしも一意に定まらない。しかし、 $A^{1/2} \geq 0$ となる平方根が必ず存在し、かつ一意に定まる。また $A > 0$ ならば、 $(A^{1/2})^{-1} = (A^{-1})^{1/2}$ が成り立ち、そのような行列を $A^{-1/2}$ と表す。

さて、この記述から $\hat{\mathbf{R}}\hat{\mathbf{V}}\hat{\mathbf{R}}^\top \geq 0$ での平方根の存在が保証できた。しかしこのままでは平方根は一意に定まらない。そこで教科書の記述を見てみると、 $(\hat{\mathbf{R}}\hat{\mathbf{V}}\hat{\mathbf{R}}^\top)^{-1/2}$ を利用していることに気付く。これは、 $\hat{\mathbf{R}}\hat{\mathbf{V}}\hat{\mathbf{R}}^\top > 0$ を暗黙に仮定することによってはじめて得られる記法である。これには追加的な仮定が必要だが、この理由と具体的内容は以下の議論によって示される:

$$\begin{aligned} \text{rank}(\hat{\mathbf{R}}\hat{\mathbf{V}}\hat{\mathbf{R}}^\top) &= \text{rank}((\hat{\mathbf{R}}\hat{\mathbf{V}})\hat{\mathbf{R}}^\top) \\ &\leq \text{rank}(\hat{\mathbf{R}}\hat{\mathbf{V}}) \leq \text{rank}(\hat{\mathbf{R}}) \quad (\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}) \\ &\implies \text{rank}(\hat{\mathbf{R}}\hat{\mathbf{V}}\hat{\mathbf{R}}^\top) \leq \text{rank}(\hat{\mathbf{R}}). \end{aligned}$$

このような関係が成り立つため、 $\hat{\mathbf{R}}$ の行フルランクは $\hat{\mathbf{R}}\hat{\mathbf{V}}\hat{\mathbf{R}}^\top$ のフルランクの必要条件である。つぎに、フルランクと正値定符号行列との関係性を確認しておこう。正定値行列は正則、つまりフルランクである。対偶をとれば、フルランクでなければ正定値行列ではない。よって、 $\hat{\mathbf{R}}$ は $\hat{\mathbf{R}}\hat{\mathbf{V}}\hat{\mathbf{R}}^\top$ が正定値行列であることの必要条件である。 $\hat{\mathbf{V}} > 0$ かつ $\text{rank}(\hat{\mathbf{R}}) = q$ で行フルランクを仮定する。このとき、次の同値条件が成り立つ:

$$\text{rank}(\hat{\mathbf{R}}) = q \iff \ker(\hat{\mathbf{R}}^\top) = \{0\} \iff \hat{\mathbf{R}}^\top x = 0 \text{ かつ } x \in \mathbb{R}^q \text{ が非ゼロなら矛盾.}$$

ここでは次数定理を利用した。すなわち、任意の $x \in \mathbb{R}^q$, $x \neq 0$ に対して $\hat{\mathbf{R}}^\top x \neq 0$ ($\in \mathbb{R}^{k+1}$) である。 $\hat{\mathbf{R}}\hat{\mathbf{V}}\hat{\mathbf{R}}^\top$ が正定値行列であることを示すためには、任意の $x \neq 0$ に対して以下を示せばよい:

$$x^\top (\hat{\mathbf{R}}\hat{\mathbf{V}}\hat{\mathbf{R}}^\top) x > 0$$

ここで仮定より、 $\hat{\mathbf{R}}^\top x \neq 0$ かつ $\hat{\mathbf{V}} > 0$ なので、

$$x^\top (\hat{\mathbf{R}}\hat{\mathbf{V}}\hat{\mathbf{R}}^\top) x = (\hat{\mathbf{R}}^\top x)^\top \hat{\mathbf{V}} (\hat{\mathbf{R}}^\top x) > 0, \quad \forall x \in \mathbb{R}^q, x \neq 0.$$

すなわち $\hat{\mathbf{R}}\hat{\mathbf{V}}\hat{\mathbf{R}}^\top$ は正値定符号行列である。以上により、 $\text{rank}(\hat{\mathbf{R}}) = q$ かつ $\hat{\mathbf{V}} > 0$ のもとで $\hat{\mathbf{R}}\hat{\mathbf{V}}\hat{\mathbf{R}}^\top$ が正値定符号行列であることが分かり、特に前者はこの必要条件であることが分かった。^{*6}

^{*5} 詳細は『はじめての最適化』などを見よ。この議論は Theorem 1.36, Theorem 2.14 によって成立する。

^{*6} 行列の積のランクと、正定値行列の正則性をここでは利用したが、正式な確認は付録で記載する。

Wald 統計量の計算で暗黙に置かれた仮定

1. $\mathbf{r}(\beta)$ の $\hat{\beta}$ 近傍での微分可能性とそれに伴う連続性
2. $\hat{\mathbf{R}}$ は行フルランク, ざっくりいえば, 制約を $\hat{\beta}$ まわりで特徴づけるうえでムダな制約はない
3. $\hat{\mathbf{V}}$ は正定値行列

■Wald 統計量の漸近分布 上で挙げた仮定は成り立つものとして, 更に議論を進めよう. 以下は確認した:

$$\sqrt{n} \cdot \mathbf{r}(\hat{\beta}) \xrightarrow{d} \mathcal{N}_q(\mathbf{0}, \mathbf{RVR}^\top), \quad \hat{\mathbf{R}}\hat{\mathbf{V}}\hat{\mathbf{R}}^\top \xrightarrow{p} \mathbf{RVR}^\top$$

$(\hat{\mathbf{R}}\hat{\mathbf{V}}\hat{\mathbf{R}}^\top)^{-1/2} \xrightarrow{p} (\mathbf{RVR}^\top)^{-1/2}$ (非自明, 行列版 CMT?) で, ベクトル版の Slutsky の定理を用いて:

$$\begin{aligned} \sqrt{n}(\hat{\mathbf{R}}\hat{\mathbf{V}}\hat{\mathbf{R}}^\top)^{-1/2}\mathbf{r}(\hat{\beta}) &\xrightarrow{d} (\mathbf{RVR}^\top)^{-1/2}\mathcal{N}_q(\mathbf{0}, \mathbf{RVR}^\top) \\ &= \mathcal{N}_q(\mathbf{0}, (\mathbf{RVR}^\top)^{-1/2}\mathbf{RVR}^\top(\mathbf{RVR}^\top)^{-1/2}) \\ &= \mathcal{N}_q(\mathbf{0}, (\mathbf{RVR}^\top)^{-1/2}(\mathbf{RVR}^\top)^{1/2}(\mathbf{RVR}^\top)^{1/2}(\mathbf{RVR}^\top)^{-1/2}) \\ &= \mathcal{N}_q(\mathbf{0}, \mathbf{I}_q \mathbf{I}_q) = \mathcal{N}_q(\mathbf{0}, \mathbf{I}_q) \end{aligned}$$

ここでは行列 \mathbf{RVR}^\top , $(\mathbf{RVR}^\top)^{-1/2}$ の対称性を利用した. 標準正規分布に従うことが分かったため:

$$\begin{aligned} W_n &= n \cdot \mathbf{r}(\hat{\beta})^\top \cdot (\hat{\mathbf{R}}\hat{\mathbf{V}}\hat{\mathbf{R}}^\top)^{-1} \cdot \mathbf{r}(\hat{\beta}) \\ &= \sqrt{n} \cdot \mathbf{r}(\hat{\beta})^\top \cdot (\mathbf{RVR}^\top)^{-1/2} \cdot \sqrt{n} \cdot (\mathbf{RVR}^\top)^{-1/2} \cdot \mathbf{r}(\hat{\beta}) \\ &= [\sqrt{n} \cdot (\mathbf{RVR}^\top)^{-1/2} \cdot \mathbf{r}(\hat{\beta})]^\top \cdot [\sqrt{n} \cdot (\mathbf{RVR}^\top)^{-1/2} \cdot \mathbf{r}(\hat{\beta})] \xrightarrow{d} \chi_q^2 \end{aligned}$$

故に, Wald 統計量は自由度 q の χ^2 分布を漸近分布に持つ. $q = 1$ のとき, $r(\beta)$ の漸近分布の共分散行列 \mathbf{RVR}^\top は漸近分散に対応しており, $SE[r(\hat{\beta})] \equiv \sqrt{\mathbf{RVR}^\top/n}$ であるため, 以下のように記載できる:

$$W_n = \left(\frac{r(\hat{\beta})}{SE[r(\hat{\beta})]} \right)^2 = t_n^2$$

■p.62 線形制約での Wald 統計量 ここまでは帰無仮説 $H_0: \mathbf{r}(\beta) = \mathbf{0}$, where $\mathbf{r}: \mathbb{R}^{k+1} \rightarrow \mathbb{R}^q$ を考えてきたが, ここでは関数形を $H_0: \mathbf{R}\beta = \mathbf{c}$ と線形特定化して考えている. 係数行列 \mathbf{R} はその表記文字の今までの対応関係を見れば想像できるが, 左辺のベクトル関数を β^\top に関して微分したヤコビ行列と一致する.*7 故に検定統計量の構築にも上記と全く同様の手順を踏めばよいため, (4.11) 式と, 同様の漸近分布を得られる.

■p.63 F 統計量 導出は上級計量の Ch.5, p.10, 11 に詳しいが, 練習のため確認しておこう. 帰無仮説 $H_0: \mathbf{R}\beta = \mathbf{c}$ 検定のため, まず制約付き OLSE $\tilde{\beta}$ を未定乗数法で導出する. 残差平方和の最小化問題:

$$\tilde{\beta} \equiv \arg \min_{\mathbf{b}} SSR_r = (\mathbf{Y} - \mathbf{Xb})^\top (\mathbf{Y} - \mathbf{Xb}) \quad \text{s.t. } \mathbf{Rb} = \mathbf{c}$$

より, ラグランジュ関数を定義し,

$$\begin{aligned} \mathcal{L}(\mathbf{b}, \boldsymbol{\lambda}) &= (\mathbf{Y} - \mathbf{Xb})^\top (\mathbf{Y} - \mathbf{Xb}) + 2\boldsymbol{\lambda}^\top (\mathbf{Rb} - \mathbf{c}) \\ &= \mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{Xb} - \mathbf{b}^\top \mathbf{X}^\top \mathbf{Y} + \mathbf{b}^\top \mathbf{X}^\top \mathbf{Xb} + 2\boldsymbol{\lambda}^\top \mathbf{Rb} - 2\boldsymbol{\lambda}^\top \mathbf{c} \\ &= \mathbf{Y}^\top \mathbf{Y} - 2\mathbf{Y}^\top \mathbf{Xb} + \mathbf{b}^\top \mathbf{X}^\top \mathbf{Xb} + 2\boldsymbol{\lambda}^\top \mathbf{Rb} - 2\boldsymbol{\lambda}^\top \mathbf{c} \quad (\because \mathbf{Y}^\top \mathbf{Xb} : \text{scalar}) \end{aligned}$$

*7 先述の $\mathbf{R}\beta = \mathbf{0}$ は制約集合 $\mathbf{r}(\beta) = \mathbf{0}$ の接平面と考えることが出来るため, 線形制約の場合は係数行列と一致するのも頷ける.

上のように OLSE の導出と同じ具合に変形する。\$\boldsymbol{\lambda}\$ は \$q \times 1\$ ラグランジュ乗数ベクトル。FONC は:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{b}} : \mathbf{X}^\top \mathbf{X} \mathbf{b} = \mathbf{X}^\top \mathbf{Y} - \mathbf{R}^\top \boldsymbol{\lambda}, \quad \frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}} : \mathbf{R} \mathbf{b} = \mathbf{c}.$$

通常、非制約推定量は \$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}\$ であり、FONC は以下のように整理できる:

$$\begin{aligned} \mathbf{X}^\top \mathbf{X} \tilde{\boldsymbol{\beta}} &= (\mathbf{X}^\top \mathbf{X}) \hat{\boldsymbol{\beta}} - \mathbf{R}^\top \boldsymbol{\lambda} \implies \tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \boldsymbol{\lambda}, \\ \mathbf{R} \hat{\boldsymbol{\beta}} - \mathbf{R} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \boldsymbol{\lambda} &= \mathbf{c}, \\ \boldsymbol{\lambda} &= [\mathbf{R} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} (\mathbf{R} \hat{\boldsymbol{\beta}} - \mathbf{c}). \end{aligned}$$

これを \$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \boldsymbol{\lambda}\$ に戻すと、制約付き最小二乗推定量は以下の通り:

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top [\mathbf{R} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} (\mathbf{R} \hat{\boldsymbol{\beta}} - \mathbf{c}).$$

次に、制約付きモデルと非制約モデルの残差平方和の差を導く。非制約モデルの残差は \$\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}\$, 制約付きモデルの残差は \$\tilde{\boldsymbol{\varepsilon}} = \mathbf{Y} - \mathbf{X} \tilde{\boldsymbol{\beta}}\$ とおく。すると:

$$\begin{aligned} \tilde{\boldsymbol{\varepsilon}} &= \mathbf{Y} - \mathbf{X} [\hat{\boldsymbol{\beta}} - (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top [\mathbf{R} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} (\mathbf{R} \hat{\boldsymbol{\beta}} - \mathbf{c})] \\ &= \hat{\boldsymbol{\varepsilon}} + \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top [\mathbf{R} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} (\mathbf{R} \hat{\boldsymbol{\beta}} - \mathbf{c}), \end{aligned}$$

ここで制約付きモデルの残差平方和は:

$$\begin{aligned} \text{SSR}_r &= \tilde{\boldsymbol{\varepsilon}}^\top \tilde{\boldsymbol{\varepsilon}} = [\hat{\boldsymbol{\varepsilon}} + \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \mathcal{M}]^\top [\hat{\boldsymbol{\varepsilon}} + \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \mathcal{M}], \\ \text{where } \mathcal{M} &\equiv [\mathbf{R} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} (\mathbf{R} \hat{\boldsymbol{\beta}} - \mathbf{c}). \end{aligned}$$

展開して \$\mathbf{X}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top = \mathbf{R} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top\$ および \$\mathbf{X}^\top \hat{\boldsymbol{\varepsilon}} = \mathbf{0}\$ を用いると,

$$\begin{aligned} \text{SSR}_r &= \hat{\boldsymbol{\varepsilon}}^\top \hat{\boldsymbol{\varepsilon}} + 2 \hat{\boldsymbol{\varepsilon}}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \mathcal{M} + \mathcal{M}^\top \mathbf{R} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \mathcal{M} \\ &= \hat{\boldsymbol{\varepsilon}}^\top \hat{\boldsymbol{\varepsilon}} + \mathcal{M}^\top \mathbf{R} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \mathcal{M}, \end{aligned}$$

第2項はスカラーで、\$\mathbf{X}^\top \hat{\boldsymbol{\varepsilon}} = \mathbf{0}\$ により消える。したがって制約付きと非制約の残差平方和の差は:

$$\text{SSR}_r - \text{SSR}_u = \mathcal{M}^\top \mathbf{R} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \mathcal{M} = (\mathbf{R} \hat{\boldsymbol{\beta}} - \mathbf{c})^\top [\mathbf{R} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} (\mathbf{R} \hat{\boldsymbol{\beta}} - \mathbf{c}).$$

この差を使って F 統計量を定義し、以下のように変形する:

$$\begin{aligned} F_n &\equiv \frac{(\text{SSR}_r - \text{SSR}_u)/q}{\text{SSR}_u/(n-k-1)} = \frac{(\mathbf{R} \hat{\boldsymbol{\beta}} - \mathbf{c})^\top [\mathbf{R} s^2 (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} (\mathbf{R} \hat{\boldsymbol{\beta}} - \mathbf{c})}{q} \\ &= \frac{n (\mathbf{R} \hat{\boldsymbol{\beta}} - \mathbf{c})^\top [\mathbf{R} \tilde{\mathbf{V}} \mathbf{R}^\top]^{-1} (\mathbf{R} \hat{\boldsymbol{\beta}} - \mathbf{c})}{q} \quad \text{where: } \tilde{\mathbf{V}} = s^2 \left(\frac{1}{n} \mathbf{X}^\top \mathbf{X} \right)^{-1} \end{aligned}$$

ここで \$\text{SSR}_u/(n-k-1) = \sum_{i=1}^n \hat{u}_i^2/(n-k-1) \equiv s^2\$, \$\tilde{\mathbf{V}}\$ は均一分散下の推定量だった。均一分散では:

$$F_n \xrightarrow{d} F(q, \infty).$$

不均一分散では一致性の崩れから Slutsky の定理を通じて分布収束も崩れるため、White 推定量で代用し:

$$F_n = W_n/q = n (\mathbf{R} \hat{\boldsymbol{\beta}} - \mathbf{c})^\top [\mathbf{R} \hat{\mathbf{V}} \mathbf{R}^\top]^{-1} (\mathbf{R} \hat{\boldsymbol{\beta}} - \mathbf{c})/q, \quad F_n \cdot q \xrightarrow{d} \chi_J^2.$$

と、結果的に推定対象は線形制約下での Wald 統計量を制約式の本数で割ったものと同じになる。この点で、不均一分散にもともと頑健な Wald 統計量の方を扱うのが良いと考えられる。

■複合仮説検定のまとめ 最後に、長ったらしい議論から、実践で注意しておくべきことを抜き出しておこう。

教訓: 複合仮説検定

1. 検定のため、制約式 $\mathbf{r}(\beta)$ には一定の条件がある.
 - (a) $\mathbf{r}(\beta)$ の $\hat{\beta}$ 近傍での微分可能性とそれに伴う連続性
 - (b) $\hat{\mathbf{R}}$ は行フルランク, ざっくりいえば, 制約を $\hat{\beta}$ まわりで特徴づけるうえでムダな制約はない
 - (c) $\hat{\mathbf{V}}$ は正定値行列
2. Wald 統計量は単一制約の下では t 統計量の二乗に一致する (t 統計量の一般化).
3. F 統計量は均一分散を仮定すれば $\xrightarrow{d} F_{q,\infty}$, 不均一分散なら $= W_n/q$.
4. 実践的には不均一分散に頑健な Wald 統計量をつねに扱えばよい.

4.4 漸近効率性と GLS 推定量

■p.66 切片無し単回帰の OLSE 漸近分布 単回帰モデル (切片なし) を考える:

$$y_i = \beta x_i + \varepsilon_i, \quad i = 1, \dots, n,$$

ここで誤差項は $\mathbb{E}[\varepsilon_i | x_i] = 0$, $\text{Var}(\varepsilon_i | x_i) = \mathbb{E}[\varepsilon_i^2 | x_i] = \sigma^2(x_i) > 0$. OLS 推定量は

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} = \beta + \frac{\sum_{i=1}^n x_i \varepsilon_i}{\sum_{i=1}^n x_i^2}.$$

これに対して、高次モーメントの有限性が守られる範囲での ε について成り立つ:

$$\frac{1}{n} \sum_{i=1}^n x_i^2 \xrightarrow{p} E[x^2], \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i \varepsilon_i \xrightarrow{d} N(0, E[x^2 \sigma^2(x)])$$

を用いると,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N\left(0, \frac{E[x^2 \sigma^2(x)]}{(E[x^2])^2}\right).$$

これに具体例を当てはめれば、教科書に記載された OLSE の漸近分布と一致する。

■p.67 一般化最小二乗法 (GLS) 次の線形モデルを考える:

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon, \quad \text{where: } \mathbb{E}[\varepsilon | \mathbf{X}] = \mathbf{0}, \mathbb{E}[\varepsilon \varepsilon^\top | \mathbf{X}] = \mathbf{\Omega} = \text{diag}(\sigma^2(\mathbf{X}_1), \dots, \sigma^2(\mathbf{X}_n)) > 0.$$

このとき、共分散行列が対称行列であることを利用して、GLS 推定量は次の最小化問題の解となる:

$$\begin{aligned} \min_{\mathbf{b}} (\mathbf{Y} - \mathbf{Xb})^\top \mathbf{\Omega}^{-1} (\mathbf{Y} - \mathbf{Xb}) &= \min_{\mathbf{b}} \sum_{i=1}^n (Y_i - \mathbf{X}_i^\top \mathbf{b}) \sigma^2(\mathbf{X}_i)^{-1} (Y_i - \mathbf{X}_i^\top \mathbf{b}) \\ &= \min_{\mathbf{b}} \sum_{i=1}^n \frac{1}{\sigma^2(\mathbf{X}_i)} (Y_i - \mathbf{X}_i^\top \mathbf{b})^2 = \min_{\mathbf{b}} \sum_{i=1}^n \frac{\varepsilon_i^2}{\sigma^2(\mathbf{X}_i)} \end{aligned}$$

上のように、各成分を分散で割ったのちの OLS 推定として解釈可能. スカラーと対称行列に注意して,

$$\begin{aligned} &\min_{\mathbf{b}} \mathbf{Y}^\top \mathbf{Y} - \mathbf{b}^\top \mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{Y} - \mathbf{Y}^\top \mathbf{\Omega}^{-1} \mathbf{Xb} + \mathbf{b}^\top \mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{Xb} \\ &= \min_{\mathbf{b}} \mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{\Omega}^{-1} \mathbf{Xb} - \mathbf{Y}^\top \mathbf{\Omega}^{-1} \mathbf{Xb} + \mathbf{b}^\top \mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{Xb} \end{aligned}$$

FONC は以下のように書け, GLS 推定量は以下のように得られる:

$$\mathbf{0} = 2 \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{Y} + 2(\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X}) \mathbf{b} = -2 \mathbf{X}^\top \boldsymbol{\Omega}^{-1} (\mathbf{Y} - \mathbf{X} \mathbf{b})$$

$$\boxed{\hat{\boldsymbol{\beta}}^{\text{GLS}} = (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{Y}}$$

$$\hat{\boldsymbol{\beta}}^{\text{GLS}} = \boldsymbol{\beta} + (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon} = \boldsymbol{\beta} + \left(\frac{1}{n} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X} \right)^{-1} \left(\frac{1}{n} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon} \right).$$

不偏性はこの変形に対し LIE を利用することで容易に検証できる.

■p.67 一致性 (GLS) 次に, GLS 推定量について一致性を (大まかに) 示す.

Proof. ここで次を定義する:

$$A_n^\Omega := \frac{1}{n} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X} \in \mathbb{R}^{(K+1) \times (K+1)}, \quad b_n^\Omega := \frac{1}{n} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon} \in \mathbb{R}^{(K+1)}.$$

i.i.d. 性, $\mathbb{E}[\|\mathbf{X}_i\|^2]$, $\mathbb{E}[\varepsilon_i^2]$, $\|\boldsymbol{\Omega}^{-1}\| < \infty$ から, LLN が使え:

$$A_n^\Omega \xrightarrow{p} \mathbb{E}\left[\frac{1}{\sigma^2(\mathbf{X}_i)} \mathbf{X}_i \mathbf{X}_i^\top\right] =: A^\Omega, \quad b_n^\Omega \xrightarrow{p} \mathbb{E}\left[\frac{\varepsilon_i}{\sigma^2(\mathbf{X}_i)} \mathbf{X}_i\right] = \mathbf{0},$$

ここでは対角行列の逆行列が, 元の対角成分を全て逆数にしたものであることを利用した. 行列 A^Ω は正則と仮定すると, CMT より $(A_n^\Omega)^{-1} \xrightarrow{p} (A^\Omega)^{-1}$. 以上から

$$\hat{\boldsymbol{\beta}}^{\text{GLS}} = \boldsymbol{\beta} + (A_n^\Omega)^{-1} b_n^\Omega \xrightarrow{p} \boldsymbol{\beta} + (A^\Omega)^{-1} \mathbf{0} = \boldsymbol{\beta},$$

すなわち GLS 推定量も一致的であることが示された. □

■p.67 漸近分布 (GLS) さらに漸近分布についても記述を確認しよう.

Proof. GLS 推定量について, まず以下のように整理する:

$$\sqrt{n} (\hat{\boldsymbol{\beta}}^{\text{GLS}} - \boldsymbol{\beta}) = \left(\frac{1}{n} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X} \right)^{-1} \left(\frac{1}{\sqrt{n}} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon} \right).$$

i.i.d. 性, $\mathbb{E}[\mathbf{X}_i \boldsymbol{\Omega}^{-1} \varepsilon_i] = \mathbf{0}$, $\mathbb{E}[\|\mathbf{X}_i \boldsymbol{\Omega}^{-1} \varepsilon_i\|^2] \leq \|\boldsymbol{\Omega}^{-1}\|^2 \sqrt{\mathbb{E}[\|\mathbf{X}_i\|^4]} \sqrt{\mathbb{E}[\varepsilon_i^4]} < \infty$ から, CMT より

$$\frac{1}{\sqrt{n}} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon} = \frac{1}{\sqrt{n}} \sum_i \frac{\varepsilon_i}{\sigma^2(\mathbf{X}_i)} \mathbf{X}_i \xrightarrow{d} N\left(\mathbf{0}, \mathbb{E}\left[\frac{\varepsilon_i^2}{\sigma^4(\mathbf{X}_i)} \mathbf{X}_i \mathbf{X}_i^\top\right]\right).$$

$\left(\frac{1}{n} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X}\right)^{-1} \xrightarrow{p} (\mathbb{E}[\frac{1}{\sigma^2(\mathbf{X}_i)} \mathbf{X}_i \mathbf{X}_i^\top])^{-1}$ と合わせて Slutsky の定理を適用すると:

$$\begin{aligned} \sqrt{n} (\hat{\boldsymbol{\beta}}^{\text{GLS}} - \boldsymbol{\beta}) &\xrightarrow{d} \mathcal{N}_{K+1}\left(\mathbf{0}, (\mathbb{E}[\frac{1}{\sigma^2(\mathbf{X}_i)} \mathbf{X}_i \mathbf{X}_i^\top])^{-1} \mathbb{E}[\frac{\varepsilon_i^2}{\sigma^4(\mathbf{X}_i)} \mathbf{X}_i \mathbf{X}_i^\top] (\mathbb{E}[\frac{1}{\sigma^2(\mathbf{X}_i)} \mathbf{X}_i \mathbf{X}_i^\top])^{-1}\right) \\ &\stackrel{d}{=} \mathcal{N}_{K+1}\left(\mathbf{0}, (\mathbb{E}[\frac{1}{\sigma^2(\mathbf{X}_i)} \mathbf{X}_i \mathbf{X}_i^\top])^{-1} \mathbb{E}[\frac{\mathbb{E}[\varepsilon_i^2 | \mathbf{X}_i]}{\sigma^4(\mathbf{X}_i)} \mathbf{X}_i \mathbf{X}_i^\top] (\mathbb{E}[\frac{1}{\sigma^2(\mathbf{X}_i)} \mathbf{X}_i \mathbf{X}_i^\top])^{-1}\right) \\ &\stackrel{d}{=} \mathcal{N}_{K+1}\left(\mathbf{0}, (\mathbb{E}[\frac{1}{\sigma^2(\mathbf{X}_i)} \mathbf{X}_i \mathbf{X}_i^\top])^{-1}\right) \end{aligned}$$

したがって GLS の漸近分布 (分散) は:

$$\boxed{\sqrt{n} (\hat{\boldsymbol{\beta}}^{\text{GLS}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}_{K+1}\left(\mathbf{0}, \mathbb{E}\left[\frac{1}{\sigma^2(\mathbf{X}_i)} \mathbf{X}_i \mathbf{X}_i^\top\right]^{-1}\right)}$$

と示される. □

ここでは内生性を考慮せず, 純粋な不均一分散の問題のみを扱っていることには注目すべきだろう.

■**効率性** GLSE の効率性についても言及しておこう。証明は上級計量の Ch.9, Appendix を参考とした。

Proof. まず OLS 推定量の条件付き分散を計算する。不偏性と以下の変形で結論できる：

$$\begin{aligned}\hat{\beta}^{\text{OLS}} - \beta &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} - \beta = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \varepsilon, \\ \text{Var}[\hat{\beta}^{\text{OLS}} | \mathbf{X}] &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \text{Var}[\varepsilon | \mathbf{X}] \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega} \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1}.\end{aligned}$$

ここで、 $\boldsymbol{\Omega}$ は共分散行列である。同様に GLS 推定量の条件付き分散を計算する：

$$\begin{aligned}\hat{\beta}^{\text{GLS}} - \beta &= (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \varepsilon, \\ \text{Var}[\hat{\beta}^{\text{GLS}} | \mathbf{X}] &= (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \text{Var}[\varepsilon | \mathbf{X}] \boldsymbol{\Omega}^{-1} \mathbf{X} (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \\ &= (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} \boldsymbol{\Omega}^{-1} \mathbf{X} (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} = (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1}.\end{aligned}$$

以上より差分は

$$\Delta := \text{Var}[\hat{\beta}^{\text{OLS}} | \mathbf{X}] - \text{Var}[\hat{\beta}^{\text{GLS}} | \mathbf{X}] = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega} \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} - (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1}.$$

ここで以下の恒等式：

$$\begin{aligned}(\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} &= (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} \boldsymbol{\Omega}^{-1} \mathbf{X} (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \\ &= (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1},\end{aligned}$$

を利用して、以下のような変形の結果、行列の積形式での表現を得る：

$$\begin{aligned}\Delta &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega} \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} - (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega} \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} - (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} - (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} + (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega} \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} - (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega} \boldsymbol{\Omega}^{-1} \mathbf{X} (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \\ &\quad - (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} + (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} \boldsymbol{\Omega}^{-1} \mathbf{X} (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \\ &= \{(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1}\} \boldsymbol{\Omega} \{(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1}\}^\top \\ &= \mathbf{Q} \boldsymbol{\Omega} \mathbf{Q}^\top \quad \text{where: } \mathbf{Q} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1}.\end{aligned}$$

$\boldsymbol{\Omega}^{-1}$ の存在性で我々は暗に $\boldsymbol{\Omega} > 0$ を仮定している。よって $\boldsymbol{\Omega} \neq \mathbf{I}_n$ のとき、*⁸ \mathbf{Q} が行フルランクなので*⁹ Wald 統計量と同様の議論により、 Δ は正定値行列で、GLS が OLS よりも小さい条件付分散を持つ。□

*⁸ $\boldsymbol{\Omega} = \mathbf{I}_n$ の時は零行列になるので除外している。この条件を外すならば Δ は半正定値行列になる。

*⁹ $\boldsymbol{\Omega}$ が正定値行列で、 \mathbf{X} が回帰モデルの仮定から列フルランクなだけでは、 \mathbf{Q} はフルランクであることは確かめられない。付録の定理より、 $\boldsymbol{\Omega}^{-1}$ は正則行列である。 $\mathbf{X}^\top \boldsymbol{\Omega}^{-1}$ は行列に正則行列を掛けた形であるため、これも付録の定理より \mathbf{X}^\top の行フルランク性が保存されている。 $(\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1}$ については、Wald 統計量での議論と正定値行列の性質から、正定値行列であるため、正則行列である。行列の積とランクの関係性から、第二項も行フルランクである。しかし、両方の項が行フルランクであっても、行列の差分が行フルランクとなる事は言えない。さらに議論を進めるため、 $\boldsymbol{\Omega} = \text{diag}(\omega_1^2, \dots, \omega_n^2)$ と対角行列に特定化する。ここで、

$$\mathbf{Q} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} =: \mathbf{Q}^{(1)} - \mathbf{Q}^{(2)},$$

この第 j 列を表示すると、標準基 \mathbf{e}_i を用いて、付録の計算により：

$$\begin{aligned}\mathbf{Q}_{\bullet j} &= \mathbf{Q} \mathbf{e}_j = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{e}_j - (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{e}_j \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} \sum_{i=1}^n e_{ij} (\mathbf{X}^\top)_{\bullet i} - (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \sum_{i=1}^n e_{ij} (\mathbf{X}^\top \boldsymbol{\Omega}^{-1})_{\bullet i} \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} \sum_{i=1}^n e_{ij} \mathbf{X}_i^\top - (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \sum_{i=1}^n \frac{1}{\omega_i^2} e_{ij} \mathbf{X}_i^\top.\end{aligned}$$

■p.67 漸近効率性 上の計算と同様の方法で漸近効率性についても計算が可能。

■p.68 実行可能な GLS (FGLS) ここで問題なのは、推定量に一般には観測不可能な Ω が入っている点だ。これにより、推定量は実行不可能である。ここで、 $\hat{\Omega}$ によって共分散行列を推定して実行可能な GLS を得る：

$$\hat{\beta}^{\text{FGLS}} = (\mathbf{X}^\top \hat{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\Omega}^{-1} \mathbf{Y}$$

$\hat{\Omega}$ が Ω の一致推定量ならば、Slutsky の定理から $\hat{\beta}^{\text{FGLS}}$ の漸近分布は $\hat{\beta}^{\text{GLS}}$ のものと一致する。

■p.68 GLS の弱点 条件付独立の仮定が満たされないとき、線形射影でもなくなるとの記述を確認しておこう。まず、線形射影モデルと回帰モデルの差異を確認しておこう。

回帰と線形射影

線形回帰モデル: $\mathbb{E}[Y_i | \mathbf{X}_i] = \mathbf{X}_i^\top \beta$ と実際の生成過程が線形で、 $\mathbb{E}[u_i | \mathbf{X}_i] = 0$

線形射影モデル: 生成過程に関わらない線形最良近似, $\mathbb{E}[u_i | \mathbf{X}_i] \neq 0$ だが $\mathbb{E}[\mathbf{X}_i u_i] = \mathbf{0}$, $\mathbb{E}[u_i] = 0$

GLS で考えたものから平均独立を外した、次のモデルを考える：

$$\mathbf{Y} = \mathbf{X} \delta + \mathbf{e}, \quad \text{where: } \mathbb{E}[\mathbf{e} \mathbf{e}^\top | \mathbf{X}] = \Omega = \text{diag}(\sigma^2(\mathbf{X}_1), \dots, \sigma^2(\mathbf{X}_n)) > 0.$$

このとき、母集団について考えれば、GLS 推定量は次の最小化問題の解となる：

$$\min_{\mathbf{d}} \mathbb{E}[(Y_i - \mathbf{X}_i^\top \mathbf{d}) \sigma^2(\mathbf{X}_i)^{-1} (Y_i - \mathbf{X}_i^\top \mathbf{d})] = \min_{\mathbf{d}} \mathbb{E}\left[\frac{1}{\sigma^2(\mathbf{X}_i)} (Y_i - \mathbf{X}_i^\top \mathbf{d})^2\right]$$

$$\text{FOC: } \mathbb{E}\left[\frac{-2}{\sigma^2(\mathbf{X}_i)} \mathbf{X}_i (Y_i - \mathbf{X}_i^\top \mathbf{d})\right] = \mathbb{E}\left[\frac{-2}{\sigma^2(\mathbf{X}_i)} \mathbf{X}_i e_i\right] = \mathbf{0}, \quad \therefore \mathbb{E}\left[\frac{1}{\sigma^2(\mathbf{X}_i)} \mathbf{X}_i e_i\right] = \mathbf{0}$$

ここで、線形射影モデルでの誤差項の仮定は満たされていない。推定対象は：

$$\mathbb{E}\left[\frac{1}{\sigma^2(\mathbf{X}_i)} \mathbf{X}_i Y_i\right] = \mathbb{E}\left[\frac{1}{\sigma^2(\mathbf{X}_i)} \mathbf{X}_i \mathbf{X}_i^\top\right] \delta$$

$$\delta = \mathbb{E}\left[\frac{1}{\sigma^2(\mathbf{X}_i)} \mathbf{X}_i \mathbf{X}_i^\top\right]^{-1} \mathbb{E}\left[\frac{1}{\sigma^2(\mathbf{X}_i)} \mathbf{X}_i Y_i\right]$$

これは OLS での推定対象である β と異なる。最良線形近似を知り、(標本内の) Y_i を最も当てはまり良く予測するための利用には不適切だ。

すなわち

$$\mathbf{Q}_{\bullet,j} = \sum_{i=1}^n \left[(\mathbf{X}^\top \mathbf{X})^{-1} - \frac{1}{\omega_i^2} (\mathbf{X}^\top \Omega^{-1} \mathbf{X})^{-1} \right] \mathbf{X}_i^\top e_{ij}.$$

ここで各 i に対応するスカラー行列

$$\mathbf{H}_i := (\mathbf{X}^\top \mathbf{X})^{-1} - \frac{1}{\omega_i^2} (\mathbf{X}^\top \Omega^{-1} \mathbf{X})^{-1}$$

は、不均一分散の下では $\omega_i^2 \neq \omega_{i'}^2$ となり得るため、異なる重み付き和の係数行列として一般に互いに独立に振る舞う。仮に非零ベクトル $\mathbf{y} \in \mathbb{R}^q$ が存在して $\mathbf{Q}\mathbf{y} = \mathbf{0}$ となるとすると、

$$\mathbf{0} = \sum_{i=1}^n \mathbf{H}_i (\mathbf{X}_i^\top \mathbf{y}),$$

すなわちスカラー $\alpha_i := \mathbf{X}_i^\top \mathbf{y}$ を用いて $\sum_i \alpha_i \mathbf{H}_i = \mathbf{0}$ が成り立つ。しかし、 ω_i^2 が全て等しくない限り、各 \mathbf{H}_i の零空間はほぼ常に交わらず、よってこの線形結合がゼロになるためには全ての $\alpha_i = 0$ を要請される。すなわち $\mathbf{X}_i^\top \mathbf{y} = 0$ for all i であり、 \mathbf{X} の列フルランク性より $\mathbf{y} = \mathbf{0}$ 。したがって $\text{rank}(\mathbf{Q}) = K + 1$ が (大まかに) 示された。

4.5 TSLS 推定量

■p.69 操作変数の条件 考えるのは, p.27 のものに内生変数を複数加えた以下の構造モデル:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \cdots + \beta_k X_{k,i} + \beta_{k+1} W_{1,i} + \cdots + \beta_{k+r} W_{r,i} + e_i$$

を考える. ここで, \mathbf{W} は外生変数, \mathbf{X} は内生変数で e_i と相関をもつ. 内生性を解決するために操作変数を用いた誘導形が導入される. 内生変数が一つの, 2 章のケースで考えれば, 二つの誘導型の比較で連立方程式の形に識別の問題を帰着させることが可能で, これを整理して以下にある関連性の条件が得られた.

復習: 内生変数が一つの場合の操作変数の条件

1. 関連性: 内生変数 X_i の誘導形

$$X_i = \pi_0 + \pi_1 W_{1,i} + \cdots + \pi_r W_{r,i} + \pi_{r+1} Z_{r+1,i} + \cdots + \pi_{r+m} W_{m,i} + \nu_i$$

での操作変数 \mathbf{Z} にかかる係数 $(\pi_{r+1}, \dots, \pi_{r+m})$ は零ベクトルではない.

2. 外生性: $\text{Cov}[Z_{1,i} e_i] = \cdots = \text{Cov}[Z_{m,i} e_i] = 0$

内生変数が複数の場合も, 行う議論の流れは同じだ. 構造型で用いた説明変数ベクトルを \mathbf{X}_i と, 誘導型での説明変数ベクトルを \mathbf{Z}_i と定義する.

操作変数の条件

1. 関連性: $\text{rank}(\mathbb{E}[\mathbf{Z}_i \mathbf{X}_i^\top]) = k + r + 1$
2. 外生性: $\mathbb{E}[\mathbf{Z}_i e_i] = \mathbf{0}$
3. $\text{rank}(\mathbb{E}[\mathbf{Z}_i \mathbf{Z}_i^\top]) = r + m + 1$

誘導モデルでの建付けから, 一般の外生性は内生変数が一つの場合の外生性と言っていることは同じ.

■p.70 階数条件 構造型の識別の必要条件は次数条件 (丁度 or 過剰識別), 必要十分条件は階数条件, との記載が 2 章であったが, 4 章ではこれを検証している. まず階数条件を確認する. 内生変数 $\mathbf{X}_{1,i}$ の誘導型は:

$$\mathbf{X}_{1,i} = \Pi_0 + \Pi_1 \mathbf{W}_i + \Pi_2 \mathbf{Z}_{1,i} + \nu_i$$

であり, $\mathbf{X}_i^\top = (1, \mathbf{X}_{1,i}^\top, \mathbf{W}_i^\top)$. これを構造モデルに代入すれば:

$$\begin{aligned} Y_i &= \mathbf{X}_i^\top \boldsymbol{\beta} + e_i = (1, [\Pi_0 + \Pi_1 \mathbf{W}_i + \Pi_2 \mathbf{Z}_{1,i} + \nu_i]^\top, \mathbf{W}_i^\top) \boldsymbol{\beta} + e_i \\ &= \beta_0 + \sum_{j=1}^k \beta_j (\pi_{0,j} + \pi_{1,j}^\top \mathbf{W}_i + \pi_{2,j}^\top \mathbf{Z}_{1,i} + \nu_{j,i}) + \sum_{l=1}^r \beta_{k+l} W_{l,i} + e_i \\ &= \gamma_0 + \gamma_1^\top \mathbf{W}_i + \gamma_2^\top \mathbf{Z}_{1,i} + \eta_i = \mathbf{Z}_i^\top \boldsymbol{\gamma} + \eta_i. \end{aligned}$$

を得る. $\mathbf{Z}_i^\top = (1, \mathbf{W}_i^\top, \mathbf{Z}_{1,i}^\top)$, $\eta_i = \sum_{j=1}^k \beta_j \nu_{j,i} + e_i$. ここで, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1^\top, \gamma_2^\top)^\top$ は次の通り:

$$\gamma_0 = \beta_0 + \sum_{j=1}^k \beta_j \pi_{0,j}, \quad \gamma_1 = \sum_{j=1}^k \beta_j \pi_{1,j} + \begin{pmatrix} \beta_{k+1} \\ \vdots \\ \beta_{k+r} \end{pmatrix}, \quad \gamma_2 = \sum_{j=1}^k \beta_j \pi_{2,j}.$$

ここまでの整理を行列・ベクトルの積で書くと

$$\mathbf{Z}_i = \begin{pmatrix} 1 \\ \mathbf{W}_i \\ \mathbf{Z}_{1,i} \end{pmatrix} \in \mathbb{R}^{1+r+m}, \quad \boldsymbol{\beta} \in \mathbb{R}^{1+k+r}, \quad \boldsymbol{\gamma} = \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{pmatrix} \in \mathbb{R}^{1+r+m}.$$

これらを用いて、線形写像 $f: \mathbb{R}^{1+k+r} \mapsto \mathbb{R}^{1+r+m}$ を以下のように表現できる:

$$\boldsymbol{\gamma} = \underbrace{\begin{pmatrix} 1 & \pi_{0,1} & \pi_{0,2} & \cdots & \pi_{0,k} & 0 & \cdots & 0 \\ \mathbf{0} & \boldsymbol{\pi}_{1,1} & \boldsymbol{\pi}_{1,2} & \cdots & \boldsymbol{\pi}_{1,k} & \mathbf{I}_r & & \\ \mathbf{0} & \boldsymbol{\pi}_{2,1} & \boldsymbol{\pi}_{2,2} & \cdots & \boldsymbol{\pi}_{2,k} & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix}}_{=:M} \boldsymbol{\beta}, \text{ where: } M \in \mathbb{R}^{(1+r+m) \times (1+k+r)}.$$

未知の $\boldsymbol{\beta}$ を一意に推定するには、つまり、線形写像 $\boldsymbol{\beta} \mapsto \boldsymbol{\gamma}$ が一意の逆写像を持つには:

$$\text{rank}(M) = 1 + r + m \implies 1 + k + r \leq 1 + r + m,$$

すなわち余剰条件 $m \geq k$ と合わせて

$$\text{rank}(M) = 1 + k + r$$

が必要十分である。この条件こそが、第一段階の係数行列 $\boldsymbol{\Pi}_2$ の列フルランク

$$\text{rank}(\boldsymbol{\Pi}_2) = k \iff \text{rank}(\mathbb{E}[\mathbf{Z}_i \mathbf{X}_i^\top]) = k + r + 1$$

として知られる階数条件であり、構造パラメータ $\boldsymbol{\beta}$ の識別可能性を保証する。これは

$$\text{rank}(\mathbb{E}[\mathbf{Z}_i \mathbf{X}_i^\top]) = k + r + 1$$

が必要十分条件となる。一方、このモーメント行列は第一段階の係数行列 $\boldsymbol{\Pi}_2$ の列空間を反映しており、 $\text{rank}(\boldsymbol{\Pi}_2) = k$ と同値であるため、階数条件は

$$\text{rank}(\boldsymbol{\Pi}_2) = k \iff \text{rank}(\mathbb{E}[\mathbf{Z}_i \mathbf{X}_i^\top]) = k + r + 1$$

となる。なぜなら、もし $\text{rank}(\boldsymbol{\Pi}_2) = k$ であれば $\mathbf{Z}_{1,i}$ のすべての成分が独立に各内生変数に寄与し、結果として $\mathbb{E}[\mathbf{Z}_i \mathbf{X}_i^\top]$ の列空間も $k + r + 1$ 次元になるからである。逆にこの階数が満たされない場合、一部の内生変数が操作変数で説明不能となり、構造パラメータは識別不能となるからである。

付録 A 計算上のテクニック

■**畳み込み積分 (Convolution)** 畳み込みとは、和が一定となるようなものをかけて足し合わせるという操作のことである。これを積分で利用したのが畳み込み積分である。

$$h(x) = \int_{-\infty}^{\infty} f(t)g(x-t)dt$$

右辺は x の関数。この $h(x)$ を $f(x)$ と $g(x)$ の畳み込み積分 (あるいは単に畳み込み) などと呼び、 $h(x) = f(x) * g(x)$ などと表記する。確率変数 X と Y の確率分布が分かっているときに、 $Z = X + Y$ の確率分布は畳み込みで与えられる。つまり：

$$h(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx$$

と計算できる。ここで、 h, f, g はそれぞれ順に連続確率変数 $Z = X + Y, X, Y$ の確率密度関数である。

■**行列の作用による線形写像** $A = (a_{k\ell}) \in \mathbb{R}^{m \times n}$, $\mathbf{v} = (v_1, \dots, v_n)^\top \in \mathbb{R}^n$, $\mathbf{u} \in \mathbb{R}^n$,

$$(A\mathbf{v})_k = \sum_{\ell=1}^n a_{k\ell} v_\ell, \quad k = 1, \dots, m.$$

分配法則は、成分ごとに表示して以下のように示せる：

$$(A(\mathbf{u} + \mathbf{v}))_k = \sum_{\ell=1}^n a_{k\ell}(u_\ell + v_\ell) = \sum_{\ell=1}^n a_{k\ell}u_\ell + \sum_{\ell=1}^n a_{k\ell}v_\ell = (A\mathbf{u})_k + (A\mathbf{v})_k,$$

スカラー倍についても、同様に示せる：

$$(A(c\mathbf{u}))_k = \sum_{\ell=1}^n a_{k\ell}(cu_\ell) = \sum_{\ell=1}^n c(a_{k\ell}u_\ell) = (c(A\mathbf{u}))_k,$$

よって、以下の線形性が成り立つ：

$$A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}, \quad A(c\mathbf{u}) = cA\mathbf{u},$$

まず標準基底による展開を確認する、長さ n の標準基底を $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ とすると、任意の $\mathbf{v} \in \mathbb{R}^n$ は $\mathbf{v} = \sum_{i=1}^n v_i \mathbf{e}_i$ と表せる。これに線形性を適用して

$$A\mathbf{v} = A\left(\sum_{i=1}^n v_i \mathbf{e}_i\right) = \sum_{i=1}^n v_i A\mathbf{e}_i = \sum_{i=1}^n v_i A_{\bullet i}.$$

ここで $A\mathbf{e}_i$ は A の第 i 列ベクトル $A_{\bullet i}$. 列ベクトル展開の結論として、特に $\mathbf{v} = \mathbf{e}_j$ のとき、

$$A\mathbf{e}_j = \sum_{i=1}^n e_{ij} A_{\bullet i}.$$

付録 B 重要な定理等

■**デルタ法** 本文でも述べた通り、標本平均の関数をうまく分布収束に持ち込める手法である。

Prop: デルタ法 (delta method), 1 変数版

$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ とする。ただし、 $\hat{\theta}_n$ は θ の一致推定量であり、 $\sigma^2 > 0$ は有限の定数とする。
関数 $g: \mathbb{R} \rightarrow \mathbb{R}$ が θ において 1 階微分可能で、 $g'(\theta) \neq 0$ のとき、

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \xrightarrow{d} \mathcal{N}(0, [g'(\theta)]^2 \sigma^2)$$

Prop: デルタ法 (delta method), 多変数版

$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}_k(\mathbf{0}, \mathbf{V})$ として、ここで関数 $g: \mathbb{R}^k \rightarrow \mathbb{R}$ が θ において微分可能であり、 \mathbf{V} は共分散行列。かつ $\nabla g(\theta) \neq \mathbf{0}$ (g は点 θ において正則) とする。このとき、

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \xrightarrow{d} \mathcal{N}(0, \nabla g(\theta)^\top \mathbf{V} \nabla g(\theta))$$

Prop: デルタ法 (delta method)

$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}_k(\mathbf{0}, \mathbf{V})$ とする。 $\hat{\theta}_n \in \mathbb{R}^k$, $\mathbf{V} \in \mathbb{R}^{k \times k}$ は共分散行列。ベクトル関数 $\mathbf{h}: \mathbb{R}^k \rightarrow \mathbb{R}^m$ が点 θ において微分可能で、ヤコビ行列:

$$\mathbf{H} := J_{\mathbf{h}}(\theta) = \left[\frac{\partial h_i}{\partial \theta_j^\top}(\theta) \right]_{1 \leq i \leq m, 1 \leq j \leq k} \in \mathbb{R}^{m \times k}$$

が存在し、かつ零行列でないとき、

$$\sqrt{n}(\mathbf{h}(\hat{\theta}_n) - \mathbf{h}(\theta)) \xrightarrow{d} \mathcal{N}_m(\mathbf{0}, \mathbf{H} \mathbf{V} \mathbf{H}^\top) \in \mathbb{R}^m$$

尚、共分散行列 $\mathbf{H} \mathbf{V} \mathbf{H}^\top$ を正値定符号行列として議論を進めたい場合には \mathbf{H} をフルランクとして議論を進める必要がある事には注意すべきだろう。実際、4.3 章ではこの仮定が途中で暗黙に課されて議論が進む。

■**行列積と rank** $\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$ が任意の 2 行列で成立する **証明**

■**正則行列と行列の積での rank** 行列の積が定義される任意の正則行列 B, C と任意の行列 A について、 $\text{rank}(BA) = \text{rank}(A)$, $\text{rank}(AC) = \text{rank}(A)$ が成立する **証明**

■**フルランク行列と行列の積での rank** 行列の積が定義される任意の列フルランク行列 B , 行フルランク行列 C と任意の行列 A について、 $\text{rank}(BA) = \text{rank}(A)$, $\text{rank}(AC) = \text{rank}(A)$ が成立する。『計量経済学のための数学』とかで扱うかも。

■**正則と正定値行列** 正定値行列は正則で逆行列も正定値行列 **証明**

参考文献

- [1] 末石 直也 (2015), 計量経済学 ミクロデータ分析へのいざない, 第 1 版, 日本評論社
- [2] 星野 匡郎, 田中 久稔, 北川 梨津 (2023), R による実証分析: 回帰分析から因果分析へ, 第 2 版, オーム社
- [3] 関口良行 (2014), はじめての最適化, 第 1 版, 近代科学社