SSE-DP-2022-3

操作変数法の理解へ：計量生物と計量経済の邂逅

国友直人

統計数理研究所

2022 年 11 月

SSE-DP-2022-3

Toward understanding the Instrumental Variables Methods in Biometrics and Econometrics

Naoto Kunitomo

The Institute of Statistical Mathematics

November 2022

(Summary)

In recent years, the statistical analysis of causal relations has been applied both in biometrics and econometrics. We first discuss the noncompliance statistical model by Angrist, Imbens and Rubin (1996, JASA) and discuss the instrumental variables methods in econometrics. We use the general linear structural equation model and the instrumental variables estimation methods. Since the least squares method is inconsistent, we discuss the Wald, LIML (limited information maximum likelihood), TSLS (two-stage least squares), and GMM (generalized moment method). We discuss some historical thoughts and future problems in statistical causal analysis.

操作変数法の理解へ：計量生物と計量経済の邂逅

(Toward understanding the Instrumental Variables Methods in Biometrics and Econometrics) [1]

国友直人[2]

2022 年 11 月 30 日

鍵言葉 (Key Words)

統計的因果推論, 臨床試験, 政策評価, Noncompliance, ATE, LATE, 操作変数法, 構造方程式, 識別性, TSLS, LIML

概要 (Summary)

因果関係 (causality) は統計科学を含め諸科学にとっては基本的かつ重要な分析対象である。計量生物と計量経済の分野ではこの間、統計的因果推論 (statistical causal inference) が盛んに応用されている。本稿ではまず Rubin (1974) に始まる反実仮想 (counter-factual) モデルと Angrist, Imbens and Rubin (1996, 略して AIR) による操作変数法 (instrumental variables method) を説明する。次に計量経済学における同時方程式と構造方程式 (structural equation) を簡単な需要関数の例を用いて説明する。一般の構造方程式を用いて統計的因果関係を解釈し、操作変数法を含めた構造方程式の統計的推定法を議論する。構造方程式の推定では OLS 法 (最小二乗法) は一致性を持たないので、操作変数法 (IV 法) としての Wald 法、LIML(制限情報最尤法, 分散比最小法)、TSLS(2 段階最小二乗法)、GMM(一般化積率法) などの長所と短所を説明する。さらに構造方程式を巡る歴史的展開を説明し、最後に計量生物と計量経済などにおける統計的因果分析のさらなる課題を展望する。

1　統計的因果推論を巡って

因果関係 (causality) の発見や理解は統計科学はもちろんのこと諸科学にとっては基本的問題であり、きわめて重要である。因果関係を巡っては

[1]未定稿, 2022 年 11 月 4 日の滋賀大学・研究会での報告の改訂稿。

[2]統計数理研究所, 〒 190-8562 東京都立川市緑町 10-3

これまで哲学をはじめ様々な分野で研究されてきているが、因果の定義に始まり、因果関係をデータから見出すにはにはどうしたら良いか、今日でもなお論争的な話題である。

そうした中でもこの間、計量生物、とりわけ疫学 (epidemilogy) ではしばらく前から統計的因果推論 (statistical causal inference) の議論を利用した多くの臨床試験が行われ、医学・薬学分野ではデータ分析に広く応用されている。統計的因果推論では様々な議論があるが、とりわけ計量生物では Rubin (1974) に始まる反実仮想 (counter-factual) の統計モデルが大きな影響を与えている。さらに近年では計量生物分野にとどまらず、社会・経済における政策評価、労働経済、教育経済、公共経済、開発経済など応用経済学の諸分野においても広く応用されている[3]。

本稿ではまず 2 節で Angrist, Imbens and Rubin (1996, 略して AIR) に始まる統計的因果推論における操作変数法の利用について説明する。次に 3 節では計量生物ではあまり知られていない計量経済において発展した同時方程式と構造方程式法の基本的議論を簡単な需要曲線と供給曲線を用いて紹介する。4 節では構造方程式を推定するための操作変数法の役割を説明、AIR (1976) の Noncompliance 統計モデルを構造方程式の特殊ケースとして解釈できることを説明する。さらに構造方程式の推定法の一般化を説明し、さらに計量生物における統計モデルを解釈する。次に 5 節, 6 節, 7 節では同時方程式モデルと構造方程式モデルを巡る歴史的展開、関連して部分識別問題や統計的因果推論の動向についてのコメントを述べる。最後に 8 節では今後の研究の方向性についてもの展望を述べる。また計量生物などではあまり知られていないと思われる操作変数法の漸近的性質については補論で簡単にまとめておく。

2　Rubin モデルと Noncompliance 問題

まず Rubin (1974) に始まる統計的因果推論における基本事項を復習しておこう。変数 Z_i $(i = 1, \cdots, n)$ を 1, 0 値をとるダミー変数として、$Z_i = 1$ なら処置群 (treatment group) に個体を割付け、$Z_i = 0$ なら対照群 (control group) に割付けるとしよう。割付けられた個体が処置 (例えば薬

[3]例えば 2019 年、2021 年におけるノーベル経済学賞の受賞者をみると国際的な意味で経済学関係者の間では統計的因果推論、政策評価 (policy evaluation) 問題への関心の強さが伺われる。基本的文献は Imbens and Rubin (2015)、日本語では岩崎 (2015), 西山 et al. (2019) などに統計的因果推論や政策評価問題の説明がある。本稿は事後的にはどちらかと云えば伝統的な econometrics からの説明になっている。

を服用) を受けたときの処置効果を $Y_i(1)$, 処置を受けなければ (あるいは偽薬, プラセボを飲む) 処置効果がなく $Y_i(0)$ という値をとるものとする。実際に第 i 番目の観測値 Y_i は処置を受けているか、そうでないかのどちらかなので

$$Y_i = Z_i Y_i(1) + (1 - Z_i) Y_i(0) \ \ (i = 1, \cdots, n) \tag{1}$$

である。この場合、第 i 番目の個体に対する処置効果は $Y_i(1) - Y_i(0)$ であるが、各個体は処置を受けるか受けないか、どちらか一方しか実行されないので、任意の i について $Y_i(1)$ と $Y_i(0)$ のいずれか片方は観測されないで、反実仮想 (counter factual) モデルと呼ばれている。この場合には平均処置効果 (average treatment effect) とは観測量が二つの状態で i.i.d. 確率変数となる場合には、ATE $= \mathbf{E}[Y_i(1) - Y_i(0)] = \mathbf{E}[Y_i|Z_i = 1] - \mathbf{E}[Y_i|Z_i = 0]$ で定義される。
計量生物では各個体を処置群と対照群にランダムに割付け、実現値を観測できると見なすことが標準的であり、その場合には ATE は

$$\hat{\beta} = \frac{\sum_{i=1}^n Z_i Y_i}{\sum_{i=1}^n Z_i} - \frac{\sum_{i=1}^n (1 - Z_i) Y_i}{\sum_{i=1}^n (1 - Z_i)} \tag{2}$$

により推定できる。ここでの状況を母数（未知の一定値）α ,β, 誤差項 u_i $(i = 1, \cdots, n)$ として線形回帰モデル

$$Y_i = \alpha + \beta Z_i + u_i \ \ (i = 1, \cdots, n) \tag{3}$$

と表現しよう。このとき誤差項 $\mathbf{E}[u_i|Z_i] = 0$ (Z_i を条件とする条件付期待値, 条件付分散は一定とする) ととれるので、$\mathbf{E}[Y_i(0)] = \alpha$ ($= \mathbf{E}[Y_i|Z_i = 0]$), $\mathbf{E}[Y_i(1)] - \mathbf{E}[Y_i(0)] = \beta$ ($= \mathbf{E}[Y_i|Z_i = 1] - \mathbf{E}[Y_i|Z_i = 0]$) となる。最小二乗法により観測量 Y_i を Z_i に回帰すると β $(= \mathbf{E}[Y_i(1) - Y_i(0)])$ の最小二乗推定量 $\hat{\beta}_{LS}$ は付論 A より (2) に一致する。したがって例えば $\sum_{i=1}^n (Z_i - \bar{Z})^2 \xrightarrow{p} \infty$ (ただし $\bar{Z} = (1/n) \sum_{i=1}^n Z_i$) が仮定できれば $n \to \infty$ のとき $\hat{\beta} - \beta \xrightarrow{p} 0$ となり、標準的な統計的推測の方法が適用できる。この状況の統計的処理は標準的であり、特に新たな問題は生じないように見える。

実際の臨床試験では処理群に割付けた被験者（および対照群に割付けられた被験者）が必ずしも処方された薬を服用するとは限らないことが起きることが報告されている。これはノンコンプライアンス（noncompliance,

非遵守）問題として知られていたが、Angrist-Imbens (1994), AIR (1996) はこの場合に操作変数法 (instrumental variables method) を利用して因果効果を推定することを提案した。実は Imbens-Angrist (1994) は社会・経済データ分析で直面する多くの問題では Noncompliance と同様の問題がかなり一般的な現象であり、その解決の手段として操作変数法の利用を提案したのである[4]。ここで有限個の割付け変数は離散変数で操作変数 Z_i、実際の処置変数 D_i を通じてのみ Y_i に影響を与える変数としよう。このとき操作変数 $g(Z_i)$ を用いた推定量を

$$\hat{\beta}_{IV} \quad = \frac{\sum_{i=1}^{n}(g(Z_i) - \bar{g}(Z))(Y_i - \bar{Y})}{\sum_{i=1}^{n}(g(Z_i) - \bar{g}(Z))(D_i - \bar{D})} \tag{4}$$

としよう。さらにダミー変数 D について、**Monotonicity(単調性) の仮定**（あらゆる z に対して $z \geq w$ なら $D_i(z)) \geq D_i(w)$ となる条件）を置こう。ここで操作変数 Z が離散的に $K+1$ 個の値 $(z_0, z_1, \cdots, z_K)$ をとるとすると、操作変数推定量の確率極限が次のように分解されることを示したのである。

$$\beta_{LATE.g} = \sum_{k=1}^{K} \lambda_k \ \beta_{z_k, z_{k-1}} \ , \tag{5}$$

ただし $P(w) = \mathbf{E}[D_i|Z_i = w]$, $\pi_k = P(Z = z_k)$, $\beta_{z_k,z_{k-1}} = \mathbf{E}[Y_i(1) - Y_i(0)|D_i(z_k) = 1, D_i(z_{k-1}) = 0]$,

$$\lambda_k = \frac{(P(z_k) - P(z_{k-1})) \sum_{l=k}^{K} \pi_l (g(z_l) - E[g(Z)])}{\sum_{m=1}^{K}(P(z_k) - P(z_{k-1})) \sum_{l=m}^{K} \pi_l (g(z_l) - E[g(Z)])}$$

で与えられる。ここで右辺に現われる各項は局所処置効果 (local treatment effects, LATE)、あるいは CATE (Compliers' average treatment effcects) と呼ばれている。

AIR (1996) では操作変数 Z と内生変数 D がともに二値変数 (0 or 1) の場合、4 つのタイプの被験者（never-taker(絶対に服用しない), complier(指示通りに行動する), defier(指示の反対行動をとる), always-taker(絶対に服用する)）が存在することを指摘、単調性の仮定は「defier (天邪鬼) [5] は標本の中には存在しない」を意味することを指摘した。この条件は実

[4]労働経済において Angrist が徴兵くじの数値が操作変数、Vietnam 戦争に従軍したか否かが右辺の内生変数、軍歴による労働賃金への影響を分析する際に操作変数を利用した問題を巡る研究が議論のきっかけであった。

[5]処理群に割り当てられると薬を服用せず、対照群に割り当てられると薬を服用するタイプの被験者を意味する。

際的にも重要な条件である。(4) において特に関数 $g(Z)=Z,\ K=1,$ $z_0=0, z_1=1$ とすると (4) の確率極限は

$$\beta_{IV}=\frac{\mathbf{E}[Y_i|Z_i=1]-\mathbf{E}[Y_i|Z_i=0]}{\mathbf{E}[D_i|Z_i=1]-\mathbf{E}[D_i|Z_i=0]} \tag{6}$$

となるが (分母がゼロとならないことを仮定)、これは AIR (1996), Proposition 1 の (12) 式に一致する。全被験者から delier を除く 3 タイプの中では never-taker と always-taker では (6) 式の分子はゼロとなる。したがって単調性の仮定の下では (6) で与えられる LATE は Complier の ATE と解釈できるのである。

ここで薬の服用効果は変数 D(1 or 0 をとるとして) に依存するはずなので (3) からの類推して母数（未知の一定値）α, β を用いて

$$Y_i=\alpha+\beta D_i+u_i \quad (i=1,\cdots,n) \tag{7}$$

により確率変数の誤差項 u_i を定めよう。割付け変数 Z_i の値 z(1 or 0) を固定して $\mathbf{E}[Y_i|Z_i=z]=\alpha+\beta\mathbf{E}[D_i|Z_i=z]$ となるようにとると、誤差項を $\mathbf{E}[u_i|Z_i=z]=0$ とすることができる。このとき β は (6) の β_{IV} に一致する。
ここで条件付期待値の条件付け事象は $Z_i=z$ としていることに注意する必要がある。処置群と対照群への割付け変数 Z_i は操作可能 (econometrics では外生的 exogenous と呼ばれる) であるが、結果として薬を服用するか否かの処置変数 D_i は観察者には直接に操作はできない (内生的 endogenous と呼ばれる) ことが重要な論点である。すなわち確率変数 Z_i $(i=1,\cdots,n)$ と確率変数 D_i $(i=1,\cdots,n)$ は一対一対応ではなく、ランダムな割付変数 Z_i(操作変数と呼ばれる) により処置変数 D_i の値がランダムに決まると見ると、説明変数 D_i は変数誤差 (errors-in-variables)、あるいは観測誤差 (measurement errors) を伴って観察されることが分かる。ここで u_i と D_i にともに Z_i に依存して確率的に変動するので付論 A の議論から $d=1$ or 0 とすると一般には $\mathbf{E}[u_i|D_i=1]\neq 0$ となる。このように一般には確率変数 u_i と D_i の共分散がゼロではないので、AIR(1996）の Noncompliance モデルでは表現 (7) は回帰方程式 (regression equation) ではなく構造方程式 (structural equation) と解釈しなければならなくなる。ここでは説明変数 D_i と誤差項 u_i の共分散がゼロではないので最小二乗法 (OLS) で Y_i を D_i に回帰、母数 α を推定すると無視できないバイアスが生じ、$n\to\infty$

のときある β^* が存在して

$$\hat{\beta}_{LS} = \quad \frac{\sum_{i=1}^{n}(D_i - \bar{D})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(D_i - \bar{D})^2} \xrightarrow{p} \beta^* \tag{8}$$

ただし $\bar{D} = (1/n)\sum_{i=1}^{n} D_i, \bar{Y} = (1/n)\sum_{i=1}^{n} Y_i,\ \ \beta^* - \beta \neq 0$ である。
したがって $\beta^* = \mathbf{E}[Y_i|D_i = 1] - \mathbf{E}[Y_i|D_i = 0]$ より (8) で与えられる OLS は構造方程式 (7) における β の一致推定量ではない[6]。ここで割付け変数と処置変数について complier, never-taker, always-taker, defier という 4 つのタイプの被験者があり得ることを思い起こそう。例えば極端ではあるが全員が確率 1 で defier の場合、処置群に割付けられた被験者 $\{Z_i = 1\} = \{D_i = 0\}$, 対照群に割り付けられた被験者 $\{Z_i = 0\} = \{D_i = 1\}$ となり計測結果は実験が目的とする処置効果の結果（ITT, Intention to Treated）と正反対となることもあり得るが、観測者 (あるいはデータを分析する統計家) にはどちらが正しいのかは分からない状況となる。一般的には（4 節で再び議論するが)、割付け変数 Z_i によって事象 $\{D_i = 1\}$ や事象 $\{D_i = 0\}$ をコントロールできないことから、AIR(1996) は被験者に対し単調性を仮定した上で LATE の推定量として (4) で与えられるワルド推定量[7] と呼ばれる操作変数法 (instrumental variables method) を利用したのである。

以上の議論を要約すると、Rubin モデルにおいて Noncompliance 問題が存在すると次のような統計的問題が生じる。
(i) 処理群と対照群をランダムに割付けても実際の服用・非服用は確率的に決まり (complier, never-taker, defier, always-taker が被験者として存在するので)、観測データそのもののみからは ATE が識別 (identify) できるとは限らない、
(ii) 実際に薬を服用した被験者グループと服用しなかった非被験者グループのデータに対して回帰分析など標準的な教科書が説明している i.i.d. 確率変数に基づく統計的方法で推定すると、LATE の推定にバイアスが生じるので操作変数法を利用する必要がある。

[6]目的変数 Y_i を処理変数 D_i に回帰して線形回帰モデル $Y_i = \alpha^* + \beta^* D_i + u_i^*$, 条件付期待値 $\mathbf{E}[u_i^*|D_i] = 0$ $(i = 1, \cdots, n)$ とすると $\alpha^* = \mathbf{E}[Y_i|D_i = 0]$ となる。ここで β^* は AT(As Treated) に対応するが、臨床試験で推定しようとする ATE とは異なる。

[7]変数誤差 (errors-in-variables) 問題を解決するために Wald (1940) により開発され方法である。変数誤差問題は本稿の議論を含め統計的多変量解析の諸問題と密接に関係するが (Anderson (2003))、議論を省略する。

これらのことは統計分析を臨床試験に応用する上では重要な意味がある。実際に多くの臨床試験では程度の差こそあれ Noncompliance 問題は存在すると考えられる。

なお Noncompliance 問題に関連する統計的問題としては、観測データから操作変数の妥当性を評価する検定法について、近年でも例えば Kitagawa (2015) が考察している。

3 Working (1927) の統計的需要関数

実は Noncompliance 問題に類似した現象は社会・経済のデータ分析ではかなり一般的に生じるのであり、厳密な意味で実験によるランダム化が困難な状況に対処する為の統計的分析法が計量経済分野では以前から開発されている。このことから Noncompliance 問題を含む Rubin モデルは様々な方向に拡張することができることが分かる。本稿では主に統計モデル上の観測される変数上に様々な変数誤差や観測誤差などのノイズが存在する場合、伝統的な生物統計とは異なり、問題を構造方程式による統計モデリングととらえることの可能性、統計分析の応用可能性について考察する。一般的な議論の前にまずミクロ経済学で古典的な例として知られている[8] Working (1927) に始まる需給均衡の例を題材に計量経済における構造方程式の意味を説明しておこう。統計的問題としてとらえると、Rubin の Noncompliance モデルと類似した反実仮想の統計的問題を扱っている、と解釈できるだろう。

ある時刻 t において市場で観察されるある魚の価格を P_t（アイスクリームだと何が異なるだろうか？）, 消費者が購入したい需要量を $Q_t^d(p)$ としよう。このとき変数の非負性、誤差項 (例えば様々な消費者、偶然性がある) を ϵ_t^d として

$$\log Q_t^d(p) = \alpha^d + \beta^d \times \log P_t^d + \epsilon_t^d \ \ (\beta^d < 0) \tag{9}$$

と表現しよう。ここで P_t^d は消費者が想定する価格水準、変数は対数変換しているが係数は弾力性と解釈できるのでしばしば利用される。ここで市場で観測されるデータ (P_t^{obs}, Q_t^{obs}) $(t = 1, \cdots, n)$ が利用可能として、消費税を新たに $100{\times}c$ パーセント導入するときの経済効果を評価する問題

[8] Imbens (2015) は Wright (2028) を説明に引用しているがここでの説明は Hayashi (2000) の 3 章により近い。

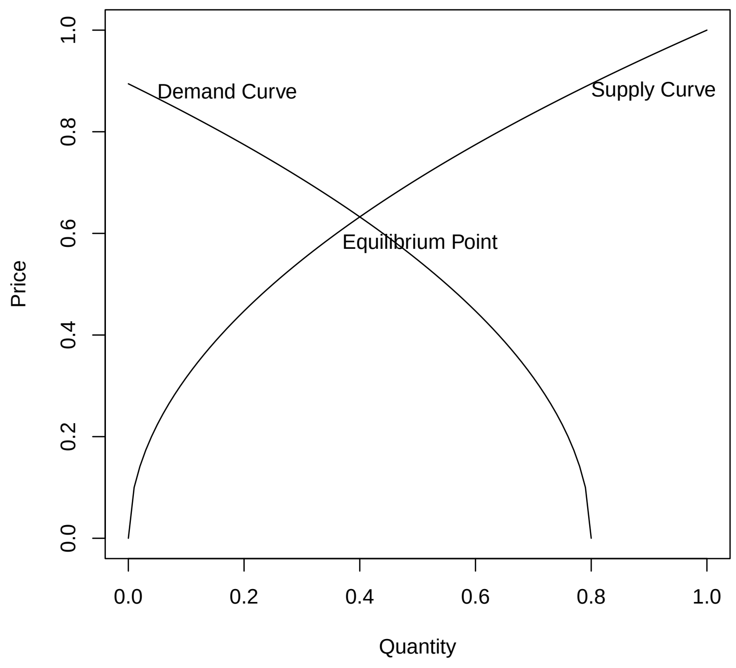

図 1 ：需要曲線 (Demand Curve) と供給曲線 (Supply Curve) の例

(図 1 では価格 (Price) と数量 (Quantity) を $[0,1] \times [0,1]$ に規準化している。市場均衡点 (Equilibrium Point) は二つの曲線の交点で表されている。)

を考えよう。ここで価格データ P を説明変数、数量データ Q を目的変数として線形回帰モデル

$$\log Q_t = \alpha + \beta \log P_t + u_t \ \ (t = 1, \cdots, T) \tag{10}$$

に最小二乗法を適用して母係数 $\hat{\alpha}, \hat{\beta}$ を推定、推定された係数に消費税率を上乗せして、消費増税の政策効果を計測することが妥当だろうか？

日本の大学で経済学を教えているごく普通の教員はこの問題について、競争的な市場で観察される価格と数量は需要と共に供給にも依存すると考えるので、回帰分析による政策評価には異議をとなえるはずである。供給関数を

$$\log Q^s_t(p) = \alpha^s + \beta^s \times \log P^s_t + \epsilon^s_t \ \ (\beta^s > 0) \tag{11}$$

と表現する。ここで $Q^s_t(p)$, P^s_t は漁業者 (or 生産者) が想定する生産と価格、ϵ^s_t は供給側の誤差項である。典型的な現代の経済学者は経済均衡 (equilibrium, 他の分野では平衡状態と表現？) 条件を考慮し、観察される市場価格と取引数量は市場で $P_t = P^d_t = P^s_t$, $Q_t = Q^d_t = Q^s_t$ が成立するように決まる、と考える。この場合には、$P_t = P^{obs}_t$ と $Q_t = Q^{obs}_t$ の連立方程式を解くと均衡価格と数量は誘導型 (reduced form) で表現

$$\begin{aligned}
\log P^{obs}_t &= \frac{\alpha^d - \alpha^s}{\beta^s - \beta^d} + \frac{\epsilon^d - \epsilon^s}{\beta^s - \beta^d}\ , \\
\log Q^{obs}_t &= \frac{\beta^s\alpha^d - \beta^d\alpha^s}{\beta^s - \beta^d} + \frac{\beta^s\epsilon^d - \beta^d\epsilon^s}{\beta^s - \beta^d}
\end{aligned}$$

となる。ここで部分均衡モデルと一般均衡モデルの差[9] を区別しなければ、消費税の効果を計測するには市場での均衡条件を $Q^d_t(P_t(c) \times (1 + c)) = Q^s_t(P_t(c))$ として構造方程式を解けば、

$$\log P^{obs}_t = \frac{\alpha^d - \alpha^s}{\beta^s - \beta^d} + \frac{\beta^d \times \log(1 + c)}{\beta^s - \beta^d} + \frac{\epsilon^d - \epsilon^s}{\beta^s - \beta^d} \tag{12}$$

となる。したがって政策評価の為には二つの構造方程式の係数に依存する $\log(1 + c)$ の係数を知る必要がある。この場合には他の条件を一定と

[9]部分均衡というミクロ経済学用語は分析対象とする市場への他の市場の影響は無視できる、市場参加者の行動は他の参加者に左右されないという条件であり、Rubin 流の SUTVA 条件の一種、と見なせることは興味深い。例えば経済学では *endogenous* (内生的) と呼んでいるが Rubin は処置効果 D_i が *not ignorable*' と呼んでいる条件に類似している。こうした計量経済と計量生物における概念の違いについては AIR(1996) を参照されたい。

すると、消費税による値上がりにより普通の消費者は (所得は一定とすると) 買い控えて需要量を減らすので、ここでの符号の仮定から平均的には観測される価格は下がる（$\beta^d/(\beta^s-\beta^d)<0$）と予想される。つまり消費税が上昇しても燃料費や賃金など生産にかかる条件には変化がないので生産者側は、同じように魚を採ろうとするが、需要が減るので結果的には市場での取引量は減少する、と予想できる。

ここで市場価格を市場数量に回帰する統計分析にでは需要関数の係数を推定できない、すなわち識別できないことは Working (1927) によりはじめて明確に指摘されたことが重要である。Working などの指摘が大きな契機となり、需要関数と供給関数などは構造型、構造方程式 (structural equation) と計量経済では呼ばれるようになったのである。次にここでとりあげた例で魚の供給関数は消費者とは関係のない天候ダミー変数 Z(悪天候で量に出られなければ 1, そうでなければ 0) を考えよう。このとき供給関数は

$$\log\ Q_t^s(p)=\alpha^s+\beta^s\times\log P_t^s+\gamma\times Z_t+\epsilon_t^s\ \ (\beta^s>0,\gamma<0) \tag{13}$$

と表現できるだろう。

この時、需要関数には表れない天候ダミー変数 Z が存在するので（除外制約 exclusion restriction が存在するので）変数 Z を割付け変数と見なすと、内生変数 Q_t,P_t に関する関係 (構造方程式としての需要関数 (9), 誤差項 ϵ_t^d は市場価格 P_t^{obs} と相関している) を操作変数 Z を利用した操作変数法により推定可能となる。同様に、あるとき海洋汚染の風評により消費者が魚の買い控え（仮にダミー変数 W とする）がおきると需要関数が変化するときには、変数 W を割付け変数として操作変数法により供給関数が推定できるのである。こうした事例における理論的分析、消費税効果などの議論は経済学の教科書に任せるとしても、外生 (exogenous) 変数 Z や W が存在すると元々の単純な構造方程式モデルと異なり、需要関数と供給関数という二つの構造方程式に現われる母数が識別可能 (identifiable) となることが本稿の議論では特に重要である[10]。

ここでとりあげた例の議論から次のようになことが云える。

(i) 市場経済で観測される価格や数量から無条件では需要関数や供給関数に表れる母数は識別できず、例えば回帰分析による消費増税の影響などの政策効果の評価は信頼できない。

[10] こうした同時方程式、構造方程式における識別問題 (identification problem) に関する古典的な条件の説明は例えば Theil (1971), 森棟 (1985) などにある。

(ii) 操作変数の存在が妥当な場合には構造方程式は識別可能となりうることが、その場合には統計学の教科書が説明している回帰モデルに対する最小二乗法で需要関数 (あるいは供給関数) を推定すると推定バイアスが大きく、データ数が大きい時には一致性を持たないので利用は不適切である。代わりに操作変数法を用いるべきである。

なお以上の例をとりあげた Imbens (2015) の説明では、確率的需要関数と確率的供給関数という二つの構造方程式に加えて市場の均衡条件により価格と数量が決まる同時方程式体系を扱っていることに注意しておく。（図 1 における二つの曲線の交点のみが観測されることを意味する。）この条件は自明なことではなく、経済学では不均衡 (disequilibrium) 状態を扱う考え方も (近年の米国経済学の主流とは云えないが) 古くから存在している。例えば不均衡状態の場合には市場での取引数量は $Q_t = \min\{Q_d^d, Q_t^s\}$ により決まり、例えば売れ残りは在庫できないと見なせば赤字覚悟で値下げして販売する、あるいは魚が売れ残りは在庫に回す、廃棄するなどが考えられる。（この場合には図 1 の左側の二つの需給曲線部分のみが観測されることになる。）ここで状態変数 D_t を (需要) $\geq$ (供給) なら $D_t = 1$, (供給) $\geq$ (需要) なら $D_t = 0$ とすると、観測される (observable) 変数 (P, Q)、(必ずしも観察できるとは限らない state variables) 状態変数 (P^d, P^s, Q^d, Q^s, D)、に関する統計モデルとして見ると $P_t = P^s D_t + P_t^d(1-D_t)$, $Q_t = Q_t^s D_t + Q_t^d(1-D_t)$ と表現できる。このように解釈すると、不均衡経済モデルには反実仮想部分を含むことになり、Rubin モデルのある種の拡張になっていることは興味深い[11]。

なお消費税の例をとると、消費者が需要を減らして一次的に市場で売れ残りが生じるのはということはある種の不均衡状態と解釈できる。こうした不均衡状態における価格調整過程は複雑になり得るが、例としては Kunitomo and Sato (1996) は一つの非線形時系列モデルを開発している。このようにここでとりあげた需給均衡の例での利用可能なデータは時系列であるので、内生・外生変数や誤差項を含めて時間効果を取り入れると、見かけ上は 2 節で議論したクロスセクションデータにもとづく臨床試験の例と異なることも重要な論点だろう。

[11] 観測データは 2 つの状態変数の片方に一致する統計モデルと見なすことが可能なので switching モデルと呼ばれているが、一種の反実仮想モデルである。本稿では詳しい議論は行わないが主要な論点は労働経済学における Quandt (1988) などの研究などと関係すがるが、不均衡モデルの nonparametric な推定法は十分に検討されていないと思われる。

4 構造方程式モデルと統計的因果推論

次に前節で説明した確率的需要関数の例を一般化して、Anderson-Kunitomo-Sawa (1982) の伝統的表現を用いて一つの線形構造方程式を

$$y_{1i} = \boldsymbol{\gamma}_1^{'} \boldsymbol{z}_{1i} + \boldsymbol{\beta}_2^{'} \boldsymbol{y}_{2i} + u_i \ \ (i = 1, \cdots, n), \tag{14}$$

としよう。ここで y_{1i} と $\boldsymbol{y}_{2i}$ は内生 (endogenous) 変数、スカラー変数と G_2 ベクトル変数、$\boldsymbol{z}_{1i}$ は K_1 個の (14) に含まれる外生 (exogenous) 変数、$\boldsymbol{\gamma}_1$ と $\boldsymbol{\beta}_2$ は $K_1 \times 1$,$G_2 \times 1$ の未知母数ベクトル、ここでは (簡単化の為) に $u_1, \cdots, u_n$ は互いに独立で分散均一の誤差項 $\mathbf{E}(u_i|\mathbf{z}_i) = 0$, $\mathbf{E}(u_i^2|\mathbf{z}_i) = \sigma^2$ $(i = 1, \cdots, n)$ を考える[12]。

次に構造方程式は同時方程式の一部であるとしよう。全体の内生変数 $\boldsymbol{y}_i^{'} = (y_{1i}, \boldsymbol{y}_{2i}^{'})^{'}$ を $1 + G_2$ として、同時方程式全体を線形多次元回帰モデル (誘導型モデルと呼ばれる) で表現でき

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\Pi} + \mathbf{V} \ , \tag{15}$$

とする[13]。ここで $\mathbf{Y} = (\mathbf{y}_i^{'})$ を $n \times (1+G_2)$ 内生変数行列、$\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2) = (\mathbf{z}_i^{'})$ は $n \times K$ 操作変数行列、K_1+K_2 個の操作変数 (instrumental variables) ベクトル $\mathbf{z}_i = (\mathbf{z}_{1i}^{'}, \mathbf{z}_{2i}^{'})^{'}$ である。$\mathbf{V} = (\mathbf{v}_i^{'})$ は $n \times (1 + G_2)$ 誤差行列、

$$\boldsymbol{\Pi} = \begin{pmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\pi}_{21} & \boldsymbol{\Pi}_{22} \end{pmatrix}$$

は $(K_1 + K_2) \times (1 + G_2)$ 係数行列、$(1 + G_2) \times (1 + G_2)$ の分散共分散行列

$$\mathbf{E}(\mathbf{v}_i \mathbf{v}_i^{'}) = \boldsymbol{\Omega} = \begin{bmatrix} \omega_{11} & \boldsymbol{\omega}_2^{'} \\ \boldsymbol{\omega}_2 & \boldsymbol{\Omega}_{22} \end{bmatrix}$$

は正定符号行列とする。ここで多変量回帰モデルとして表現できる (15) 式 $\mathbf{y}_i = \boldsymbol{\Pi}^{'} \mathbf{z}_i + \mathbf{v}_i$ $(i = 1, \cdots, n)$ は誘導型 (reduced form) モデルと呼ばれ

---

[12]内生変数 y_{1i} が $G_1 \times 1$ ベクトルの場合に拡張することが可能であり、部分システム法として知られている。4 節では説明の便宜上で $G_1 = 1$ としておくが、$G_1 \geq 1$ の場合については付論で簡単に言及しておいた。実際の治験では幾つかの変数について同時に情報が得られことが一般的と思われる。

[13]内生変数や外生変数にダミー変数が含まれていても統計モデルが線形であれば誤差項を離散変数にとればこうした表現が可能である。需要供給関数の例では例えば $G_2 = 1, K_1 = K_2 = 1$ であり需要関数が構造方程式、需給均衡式を解いた結果が 2 次元回帰モデルとなる。AIR(1996) による Rubin タイプの Noncompliance モデルについては以下の例 1 と例 2 で説明する。

ている。

なお構造方程式モデルは誤差項の分散不均一性などがある場合への拡張は可能である。また、計量経済分野 (econometrics) ではより一般の非線形モデルの場合についても様々な拡張が議論されているが、例えばしばらく前までの議論については Amemiya (1985) を参照されたい。

定式化の因果的解釈 :

ここで説明する構造方程式は処置効果モデルとしては次のように解釈できることに注意しよう。処置効果の場合には y_{1i} は第 i 番目の観測量 (処置群 or 対照群)、観測誤差も存在するものと仮定すると、構造方程式については次のようなコメントを述べておく。

(i) $\mathbf{y}_i = (y_{1i}, \mathbf{y}_{2i}')'$ は内生 (endogenous) 変数、$\mathbf{z}_i = (\mathbf{z}_{1i}', \mathbf{z}_{2i}')'$ は外生 (exogenous) 変数と呼ばれるが、(14) と (15) で表現される構造方程式モデルは $\mathbf{z}_i$ が与えられたという条件のもとでの $\mathbf{y}_i$ の分布が定まるという条件付モデルである。(14) 式に表れる内生変数 $\mathbf{y}_{2i}$ と誤差項 u_i は一般に独立でなく相関がある。一般の構造方程式では $\mathbf{z}_{1i}$ と $\mathbf{z}_{2i}$ の独立性は必ずしも必要ない。Rubin 流の因果推論モデルで $\mathbf{z}_i$ の一部分 $\mathbf{z}_{2i}$ がランダム割付け変数の場合であり、この場合は $\mathbf{z}_{1i}$ と $\mathbf{z}_{2i}$ は独立にとれる。

(ii) y_{1i} は第 i 個体に対する臨床試験の観察結果、z_{1i} は構造方程式に現れる操作変数である。$K_1 = 1, z_{1i} = 1$ とすれば定数項になり構造方程式は (処理効果+誤差項 or 誤差項) となる。$\mathbf{z}_{1i}$ は目的変数 Y_i に直接に影響を与える説明変数であり、例えば性別、人種、年齢といった割付へ変数に関係せず結果に対して影響する効果を表現することが可能である。

(iii) $\mathbf{y}_{2i} = D_i$（$G_2 = 1$ とした 1, 0 をとる 1 次元ダミー変数）とすると操作変数 Z に依存する内生変数 $D_i(Z)$ である。特に $D_i = 1$ のとき $y_{1i} = \boldsymbol{\gamma}_1' \boldsymbol{z}_{1i} + \beta_2 + u_i$、$D_i = 0$ のとき $y_{1i} = \boldsymbol{\gamma}_1' \boldsymbol{z}_{1i} + u_i$ となり、$K_1 = 1$, $z_{1i} = 1$, $K_2 = 1$, z_{2i} を 0,1 のダミー変数にとると AIR (1996) モデルに対応する。このとき 2×2 行列 $\mathbf{G} = (g_{ij})$ の要素を

$$
\begin{aligned}
g_{11} &= [\sum_{i=1}^{n} (g(Z_i) - \bar{g}(Z))(Y_i - \bar{Y})]^2 , \\
g_{12} &= (g_{21} =) \, [\sum_{i=1}^{n} (g(Z_i) - \bar{g}(Z))(Y_i - \bar{Y})][\sum_{i=1}^{n} (g(Z_i) - \bar{g}(Z))(D_i - \bar{D})] \\
g_{22} &= [\sum_{i=1}^{n} (g(Z_i) - \bar{g}(Z))(D_i - \bar{D})]^2
\end{aligned}
$$

とすると係数推定量は

$$\hat{\beta}_{2.IV} = \frac{g_{21}}{g_{22}} \tag{16}$$

と表現できる。(これは以下で説明するの (29) 式と (??) 式の特殊ケースとなるが、この場合は操作変数推定量 (IV)、二段階最小二乗推定量 (TSLS)、制限情報最尤推定量 (LIML) は丁度識別可能 (just-identified) なので一致する。)
また $K_1 = 1$, $z_{1i} = 1$, $K_2 = 1$, z_{2i} を $K + 1$ 個の値をとる離散確率変数とすると Imbens-Angrist (1994) のモデルに対応する。
(iv) 変数ベクトル $\mathbf{z}_{2i}$ は一つの構造方程式に表れないが $\mathbf{y}_{2i}$ を説明する為に必要な操作変数である。そこでよく除外制約 (exclusion restriction) と呼ばれているが、割付け変数（ベクトル）は処置変数 (ベクトル) $\mathbf{y}_{2i}$ を通じてのみ y_{1i} に影響することを意味している。例えば操作変数 (ベクトル)$\mathbf{z}_{2i}$ は一番目には処置の割付け変数、二番目以降には例えば医師の推奨連絡、ワクチン接種の書類、などランダム割付の際に利用される情報、またサンプリングの時の地域 (しばしば公的サンプリングでは大都市、中都市、農村部でサンプリング確率が変わる etc.)、職域なども考えられるだろう。
(v) AIR (1996) が引用している Angrist の実証例では y_{1i} は所得、$y_{2i} = D_i$ はベトナム戦争徴兵ダミー変数、z_{2i} は徴兵くじ番号ダミー変数、である。構造方程式には現われず (exclusion restriction), 内生変数 y_{1i} には直接には影響しない変数であるり、本当に影響しないか否かは完全には検定できないが、過剰識別性 (overidentifying) 検定と呼ばれる作業は可能である。
(vi) 右辺に存在する内生変数の数 $G_2 > 1$ の場合を考えることができる。以下の例 4 で示すように複数の水準の処理効果、例えばゼロ, 少量, 多量の 3 水準のの場合には $G_2 = 2$ という内生変数に対する処置効果の計測問題として処理が可能である。

以上の説明をまとめると、AIR(1996) の Noncompliance モデルは構造方程式モデルにおいて 1 個の除かれた外生変数が割付け変数、ダミー変数の場合と解釈できる。すなわち統計的因果推論と構造方程式モデルの関係を理解することで Noncompliance モデルは様々な方向に拡張可能となることが分かる。また、社会・経済などにおける統計的政策評価問題では、ランダム化比較実験の役割が限定的であり、厳密な意味でそれを実施することは極めて困難である場合にも適用可能な統計的方法が必要である。その意味では構造方程式モデルは有用と考えられる。

ここで全体の説明変数としての操作変数ベクトル $K\ (=K_1+K_2, n>2)$ $\mathbf{z}_i$ は直交条件 $\mathbf{E}[u_i \boldsymbol{z}_i] = \mathbf{0}\ (i=1,\cdots,n)$ を満足しているものとする．このとき、最初の構造方程式が意味あるときには制約条件

$$\begin{pmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\pi}_{21} & \boldsymbol{\Pi}_{22} \end{pmatrix} \begin{pmatrix} 1 \\ -\boldsymbol{\beta}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \mathbf{0} \end{pmatrix}, \tag{17}$$

$u_i = (1, -\boldsymbol{\beta}_2^{'})\mathbf{v}_i = \boldsymbol{\beta}^{'}\mathbf{v}_i$ および（分散・共分散が均一であれば）$\sigma^2 = \boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}$, $\boldsymbol{\beta}^{'} = (1, -\boldsymbol{\beta}_2^{'})$ が成り立つ。母数ベクトルの識別性の必要十分条件は $\mathrm{rank}(\boldsymbol{\Pi}_{22}) = G_2$ であり、必要条件は $K_2 \geq G_2$ なので、こうした多変量回帰モデルにおける係数行列の階数が退化する条件を仮定する必要がある、ことに注意しておこう。

ここで構造方程式モデルを利用して条件付期待値をとると、例えば二つの操作変数ベクトル $\mathbf{w}_1$ と $\mathbf{w}_2$ 対して $\mathbf{E}[y_{1i}|\mathbf{z}_{2i}=\mathbf{w}_j] = \mathbf{z}_{1i}^{'}\boldsymbol{\pi}_{11} + \mathbf{w}_j^{'}\boldsymbol{\pi}_{21}$,$(j=1,2)$, $\mathbf{E}[y_{2i}^{'}|\mathbf{z}_{2i}=\mathbf{w}_j] = \mathbf{z}_{1i}^{'}\boldsymbol{\Pi}_{12} + \mathbf{w}_j^{'}\boldsymbol{\Pi}_{22}\ (j=1,2)$ となるので、構造方程式に表れる $\mathbf{z}_{1i}$ が一定という条件と (17) より

$$\mathbf{E}[y_{1i}|\mathbf{z}_{2i}=\mathbf{w}_2] - \mathbf{E}[y_{1i}|\mathbf{z}_{2i}=\mathbf{w}_1] = \left(\mathbf{E}[\mathbf{y}_{2i}^{'}|\mathbf{z}_{2i}=\mathbf{w}_2] - \mathbf{E}[\mathbf{y}_{2i}^{'}|\mathbf{z}_{2i}=\mathbf{w}_1]\right)\boldsymbol{\beta}_2$$

となる。以上のことから次の結果が得られる。

命題 1：構造方程式モデル (14), (15) において操作変数 $\mathbf{z}_{2i}\ (i=1,\cdots)$ が確率的にベクトル列 $\mathbf{w}_j\ (j=0,1,\cdots,m)$ をとるとする。このとき内生変数 $\mathbf{y}_{1i}\ (i=1,\cdots)$ の条件付期待値について

$$\begin{aligned} & \mathbf{E}\left[\sum_{j=1}^{m}\left(\mathbf{E}[y_{1i}|\mathbf{z}_{2i}=\mathbf{w}_j] - \mathbf{E}[y_{1i}|\mathbf{z}_{2i}=\mathbf{w}_{j-1}]\right)\right] \\ = \ & \mathbf{E}\left[\sum_{j=1}^{m}\left(\mathbf{E}[\mathbf{y}_{2i}^{'}|\mathbf{z}_{2i}=\mathbf{w}_j] - \mathbf{E}[\mathbf{y}_{2i}^{'}|\mathbf{z}_{2i}=\mathbf{w}_{j-1}]\right)\right]\boldsymbol{\beta}_2 \end{aligned} \tag{18}$$

が成り立つ。特に $m=1$ なら

$$\boldsymbol{\beta}_2 = \frac{\partial(\mathbf{E}[y_{1i}|\mathbf{z}_{2i}=\mathbf{w}_1] - \mathbf{E}[y_{1i}|\mathbf{z}_{2i}=\mathbf{w}_0])}{\partial(\mathbf{E}[\mathbf{y}_{2i}^{'}|\mathbf{z}_{2i}=\mathbf{w}_1] - \mathbf{E}[\mathbf{y}_{2i}^{'}|\mathbf{z}_{2i}=\mathbf{w}_0])}$$

となる。

ここで特に $G_2 = K_1 = 1$, $\mathbf{y}_{1i} = Y_i$(左辺の内生変数)、また右辺の内生変数 $\mathbf{y}_{2i} = D_i$ および除かれた外生変数 $\mathbf{z}_{2i} = Z_i$ はともに二値 0,1 をとるダミー変数とすると

$$(\mathbf{E}[Y_i|Z_i = 1] - \mathbf{E}[Y_i|Z_i = 0]) = (\mathbf{E}[D_i|Z_i = 1] - \mathbf{E}[D_i|Z_i = 0])\,\beta_2$$

より LATE 係数 (6) が得られる。すなわち (18) 式の結果は Noncompliance モデルに対する操作変数推定の一般化とみなせるのである。しかしここでの導出には特に単調性の仮定などは利用していないが、Noncompliance モデルが構造方程式として表現可能なのか、統計モデルおよびその母数の識別性 (identification) を巡る興味深い論点を与えていると思われる。そこで再び前例を議論しよう。

例 1(再び)：Rubin (1974) 反実仮想モデル ：
Rubin (1974) モデルはダミー変数を用いた線形回帰モデルで表現されるが、退化した構造方程式とも見なすことができる。$G_1 = K_2 = 1, K_1 = 0$, $y_{1i} = Y_i, z_{2i} = \mathbf{y}_{2i} = D_i = Z_i, \mathbf{v}_{i2} = \mathbf{0}$(恒等的にゼロ) とすると (14) 式の時別な場合となる。この場合には観測誤差 (あるいは変数誤差) がないので右辺の変数 Z_i は外生的である。関心のある変数は常に観測できるとは限らず、処理群と対照群における結果変数 $Y_i(1), Y_i(0)$ は状態変数 (state variables) とみなすことができるのである。観測する変数は目的変数 Y_i と割付け変数 (処置変数でもある)Z_i、後者の Z_i は 2 値 1,0 をとるダミー変数である。

例 2(再び)：AIR (1996) モデル ：
再び AIR(1996) の Noncompliance モデルを取り上げる。$G_1 = K_1 = K_2 = 1$, $z_{1i} = 1$, $z_{2i} = Z_i$, $y_{1i} = Y_i$, $\mathbf{y}_{2i} = D_i$, $\mathbf{v}_{2i} = v_{2i}$(スカラーの確率変数) とおく。このとき (1) の構造方程式は $Y_i = \gamma_1 + \beta_2 D_i + u_i$, (15) の第 2 の方程式を

$$D_i = \pi_{12} + \pi_{22} Z_i + v_{2i} \ , \tag{19}$$

と表そう。ただし誤差項 v_i $(i = 1, \cdots, n)$ は $\mathbf{E}[v_i|Z_i] = 0$ を満たすように係数 π_{12}, π_{22} をとる必要がある。Noncompliance モデルではランダムに (econometrics 用語では外生的に) 割付け変数 Z_i、ランダムな割付けに対する個人の反応変数が（econometrics 用語では内生的に決まる変数）D_i (実際に個人が試薬を服用すれば 1 となる処置変数) となる。臨床試験の結果は、実際に試薬を服用したか否かを示す D_i= $1, 0$ のみに依存する (外生変数 z_{1i} は定数なので無視できる) ので、これが関心のある構造方程式、

目的変数 (やはり内生変数)y_{1i} となる。
ここで個人が Complier の場合は $Z_i = 1, D_i = 1$ $(v_{2i} = 1 - \pi_{12} - \pi_{22})$ or $Z_i = 0, D_i = 0$ $(v_{2i} = -\pi_{12})$ である、同様に Never-taker は $Z_i = 1, D_i = 0$ $(v_{2i} = -\pi_{12} - \pi_{22})$ or $Z_i = 0, D_i = 0$ $(v_{2i} = \pi_{12})$, Always-taker は $Z_i = 1, D_i = 1$ $(v_{2i} = 1 - \pi_{12} - \pi_{22})$ or $Z_i = 0, D_i = 1$ $(v_{2i} = 1 - \pi_{12})$, Defier は $Z_i = 1, D_i = 0$ $(v_{2i} = -\pi_{12} - \pi_{22})$ or $Z_i = 0, D_i = 1$ $(v_{2i} = 1 - \pi_{12})$, にそれぞれ対応する。各タイプの被験者確率をそれぞれ p_c, p_n, p_a, p_d とすると制約条件 $\sum_{k=c,n,a,d} p_k = 1$, $\mathbf{E}[v_{2i}|Z_i] = 0$ の条件が必要となる。
この期待値条件は簡単な計算から $\mathbf{E}[v_{2i}|Z_i = 1] = p_c + p_a - (\pi_{12} + \pi_{22}) = 0$, $\mathbf{E}[v_{i2}|Z_i = 0] = p_a + p_d - \pi_{12} = 0$ より $p_a + p_d = \pi_{12}$, $p_c + p_a = \pi_{12} + \pi_{22}$, $p_c + p_n + p_a + p_d = 1$ と表現できる。ここで誘導型 (reduced form) 係数 π_{12}, π_{22} はデータから推定可能なので

$$\text{条件 D：} \quad p_d = 0 \tag{20}$$

を課せば p_c, p_n, p_a を定めることができる。なおこの識別性条件は AIR (1996) の単調性条件と同値である。したがって、一般には制約条件なしには Noncompliance モデルは識別不能な (non-identifiable) 統計モデルであることが確認される。

例 3 ：One-sided Noncompliance モデル ：
例 2 のモデルの設定をほんの少し変更した One-sided Noncompliance モデルを考えよう。臨床試験では処理群に割付けられたとしても実際に薬 (あるいは Placebo) を服用しない場合も考えられるが、対照群に割付けた被験者は薬を服用することはできないように設定できる場合が少なくない。例えば COVIT ワクチン接種では地方自治体などからの公的通知書類を提出しないと接種はできないのが一般的である。この場合には自動的に 2 章で説明した単調性 (Monotonicity) の条件は満たされている。変数 Z と D の観測データから参加者は 3 種類（$Z_i = 1, D_i = 1$ をとりタイプ A, $Z_i = 1, D_i = 0$ をとるタイプ B, $Z_i = 0, D_i = 0$ をとるタイプ C）に分けられるが、それぞれの確率 p_A, p_B, p_C とすると $p_c = P\left(Z_i = 0, D_i = 0\right)$ はデータから定められるので、直交条件 $\mathbf{E}[v_{2i}|Z_i = 1] = 0$, $p_A + p_B + p_C = 1$ を利用すると p_A, p_B も定めることができ、このモデルの母数は識別される。この場合には操作変数法により consistent に LATE を推定することができる (この one-sided-nncompliance モデルについては Imbens-Rubin (2015)23 章により詳しい説明がある)。

本節では $G_2 \geq 1, K_2 \geq 1$ を含む様々な線形構造方程式を考察している。ここでは $G_2 = K_2 = 2, K_1 = 1$ となる一つの例のみを例示しておく。

例 4： 処置 3 水準の Noncompliance モデル ：
処置水準が 3 つ (ゼロ水準の投与, 少量水準の投与, 多量水準の投与) あり、Noncompliance 問題が存在する場合を一例として説明しよう。この場合には $G_2 = K_2 = 2,\ K_1 = 1$ として $y_{1i} = Y_i$(左辺の内生変数)、内生的ダミー変数 $\mathbf{D}_{ji}$ $(j =, 1, 2)$ （右辺の内生変数ベクトル $\mathbf{y}_{2i}^{'} = (D_{1i}, D_{2i})$ の内、$D_{1i} = 1$ は低水準の服用, $D_{2i} = 1$ は高水準の服用, $D_{1i} = 0, D_{2i} = 0$ はその他）、除かれた操作変数 $\mathbf{z}_{2i}$ には 2 個の割付ダミー変数 $Z_{21,i}, Z_{22,i}$ （$Z_{21,i} = 1$ は低水準の割付け, $Z_{22,i} = 1$ は高水準の割付け)、構造方程式は

$$Y_i = \gamma_1 + (D_{1i}, D_{2i})\boldsymbol{\beta}_2 + u_i \tag{21}$$

とすれば良い。ここで $\mathbf{E}[u_i|Z_{21,i} = z_{21}, Z_{22,i} = z_{22}] = 0$ とすると γ_1 は 0 水準の効果、$\boldsymbol{\beta}_2$ は 2×1 母数ベクトル、$(1,0)\boldsymbol{\beta}_2$ $(= \beta_{21})$ は低水準効果、$(0,1)\boldsymbol{\beta}_2$ $(= \beta_{22})$ は高水準効果をそれぞれ表す。二つの内生変数 D_{1i}, D_{2i} は二つの操作変数 (外生変数)$Z_{21,i}, Z_{22,i}$ に依存するので表現

$$D_{1i} = \pi_{11}^{(22)} + \pi_{12}^{(22)} Z_{21,i} + v_{21,i}\ , D_{2i} = \pi_{21}^{(22)} + \pi_{22}^{(22)} Z_{22,i} + v_{22,i}\ ,$$

$(\boldsymbol{\Pi}_{22} = (\pi_{ij}^{(22)})\,(i,j = 1,2)$ とすると、誤差項 $v_{2j,i}$ $(j = 1, 2)$ は例 2 と同様に条件 $\mathbf{E}[v_{2j,i}|\mathbf{z}_i] = 0$ となるようにとれる。
このとき例 2 と同様に被験者の態度を表すダミー変数 D_j $(j = 1, 2)$ について 4 タイプ (complier, never-taker, always-taker, defier) が存在すると考えると、それぞれの確率を $p_c(j), p_n(j), p_a(j), p_d(j)$ $(j = 1, 2)$ とすると、例 2 にと同様の議論より $p_g(1) = p_d(2) = 0$ が LATE を識別する十分条件となる。ただし、さらに二つの状態に対する各タイプの確率を共通 $p_c(1) = p_c(2) = p_a$, $p_n(1) = p_n(2) = p_n$ とすると（誘導型母数に）制約条件 $\pi_{11}^{(22)} = \pi_{21}^{(22)}$, $\pi_{12}^{(22)} = \pi_{22}^{(22)}$ を課す必要生じる。処置のレベル、薬の処方量 (ゼロの水準, 少量の水準, 多量の水準 etc.) により処置への態度が変化するということもありうるので、様々な統計モデルが考えられることは応用上で興味深い。

例 5： 層別標本 (Stratified Sampling) モデル ：
実際のサンプリングではしばしば様々な理由により単純ランダムサンプリングではなく、層別標本の場合なども少なくない。伝統的な統計学の標本調査論では層別サンプリングは単純なランダムサンプリングに比べると、

層内分散が小さく、層間分散が大きい場合には特に推定精度が向上することが知られている。全体のサンプルを J 個の層に分けてサンプリングを行う場合には $K_2 = J\ (J \geq 2)$ とすると構造方程式の推定において層別ランダムサンプリングを扱うことが可能となる。この場合には層の影響を考慮することなく処置群と対照群の割付けと One-Sided Noncompliance の構造方程式 $G_2 = 1$ により LATE を推定するには過剰識別 (overidentifying) モデル ($K_1 > G_1$ の場合) が適切となる。なお層別効果も同時に推定したい場合には右辺の内生変数を含む構造方程式モデルを利用する必要が生じると考えられる[14]。

こうした Noncompliance モデルにおける複数の変数がある場合の LATE の推定は、次に述べる推定量の漸近正規性の結果を利用すると、比較的簡単に実行可能となることがここでの定式化の長所と考えられる。しかし、こうした構造方程式を利用した統計的推定法はあまり見かけないようであるので、シミュレーションなどによりデータ数の影響などの統計的性質を調べる必要があるだろう。

構造方程式の推定法：
操作変数法としては AIR (1996) で利用した Wald (1940) 法はかなり特殊な方法であり、例えば $K_1 = K_2 = G_2 = 1$ の場合をより一般の場合に拡張しようとすると、かなりの困難性があるばかりか、補論 B で示すようにデータ数 n が大きい場合には次に述べる TSLS, LIML などに比べ統計的に非効率的であることが知られている。そこでより一般的な構造方程式の推定法を導入、その統計的性質を議論する[15]。
行列 $\mathbf{\Pi}_2 = (\boldsymbol{\pi}_{21}, \mathbf{\Pi}_{22})$ を $K_2 \times (1+G_2)$ 係数行列に分割、$(1+G_2)\times(1+G_2)$ 行列

$$\mathbf{G} = \mathbf{Y}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{Y} = \mathbf{P}_2'\mathbf{A}_{22.1}\mathbf{P}_2\ , \tag{22}$$

および

$$\mathbf{H} = \mathbf{Y}'\left(\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\right)\mathbf{Y}\ , \tag{23}$$

とする。ただし、記号 $\mathbf{A}_{22.1} = \mathbf{Z}_{2.1}'\mathbf{Z}_{2.1}$, $\mathbf{Z}_{2.1} = \mathbf{Z}_{2n} - \mathbf{Z}_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$, $\mathbf{P_2} =$

[14] こうした構造方程式モデルについては別の機会に考察する予定である。
[15] 詳しくは補論 B の命題 3 を参照されたい。計量経済では一般化積率法 (generalized method of moments, GMM) もしばしば利用されているが付論の命題 4 を参照されたい。例えば Hayashi (2000), 西山 et al. (2019) 付論に説明がある。

$\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{Y}$,

$$\mathbf{Z}_1 = \begin{pmatrix} \mathbf{z}_{11}^{'} \\ \vdots \\ \mathbf{z}_{1n}^{'} \end{pmatrix}, \mathbf{Z}_2 = \begin{pmatrix} \mathbf{z}_{21}^{'} \\ \vdots \\ \mathbf{z}_{2n}^{'} \end{pmatrix}, \tag{24}$$

により行列を表現、さらに

$$\mathbf{A} = \begin{pmatrix} \mathbf{Z}_1^{'} \\ \mathbf{Z}_2^{'} \end{pmatrix} (\mathbf{Z}_1, \mathbf{Z}_2) = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \tag{25}$$

は正則行列 (a.s.) とする。このとき $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}_2^{'})^{'}$ の LIML 推定量 $\hat{\boldsymbol{\beta}}_{LI}$ は次の固有方程式を満たす固有ベクトルで与えられる。

$$(\frac{1}{n}\mathbf{G} - \frac{1}{q_n}\lambda_1\mathbf{H})\hat{\boldsymbol{\beta}}_{LI} = \mathbf{0}, \tag{26}$$

ただし $q_n = n - K$ $(n > 2)$、λ_1 $(n > 2)$ は固有方程式

$$|\frac{1}{n}\mathbf{G} - l\frac{1}{q_n}\mathbf{H}| = 0 \tag{27}$$

最小固有値である。
ここで構造方程式 $y_{1i} - \boldsymbol{\beta}_2^{'}\boldsymbol{y}_{2i} = \boldsymbol{\gamma}_1^{'}\boldsymbol{z}_{1i} + u_i$ $(i = 1, \cdots, n)$ の左辺をまとめて $\mathbf{Y}\boldsymbol{\beta} = \mathbf{Z}_1\boldsymbol{\gamma}_1 + \mathbf{u}$ と表記すると、$\boldsymbol{\beta}$ を所与とする残差平方和は $R_1(\boldsymbol{\beta}) = \boldsymbol{\beta}^{'}\mathbf{Y}^{'}\bar{\mathbf{P}}_{Z_1}\mathbf{Y}\boldsymbol{\beta}$ となる。(なお $\mathbf{P}_{Z_1} = \mathbf{Z}_1(\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}\mathbf{Z}_1^{'}$, $\bar{\mathbf{P}}_{Z_1} = \mathbf{I}_n - \mathbf{P}_{Z_1}$ とした。) これに対して仮に $y_{1i} - \boldsymbol{\beta}_2^{'}\boldsymbol{y}_{2i} = \boldsymbol{\gamma}_1^{'}\boldsymbol{z}_{1i} + \boldsymbol{\gamma}_2^{'}\boldsymbol{z}_{2i} + u_i$ $(i = 1, \cdots, n)$ を回帰方程式と見ると表現 $\mathbf{Y}\boldsymbol{\beta} = \mathbf{Z}_1\boldsymbol{\gamma}_1 + \mathbf{Z}_2\boldsymbol{\gamma}_2 + \mathbf{u}$ より、残差平方和は $R_2(\boldsymbol{\beta}) = \boldsymbol{\beta}^{'}\mathbf{Y}^{'}\bar{\mathbf{P}}_Z\mathbf{Y}\boldsymbol{\beta}$ となる。(なお $\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}^{'}\mathbf{Z})^{-1}\mathbf{Z}^{'}$, $\bar{\mathbf{P}}_Z = \mathbf{I}_n - \mathbf{P}_Z$ とした。) 構造方程式の仮説は $H_0 : \boldsymbol{\gamma}_2 = \mathbf{0}$ という仮説であるから、この仮説のもと $\boldsymbol{\beta}$ を推定するには残差平方和の比率 $R(\boldsymbol{\beta}) = R_1(\boldsymbol{\beta})/R_2(\boldsymbol{\beta})$ あるいは分散比 $l = (R_1 - R_2)/R_2$ をとり

$$l(\boldsymbol{\beta}) = \frac{\boldsymbol{\beta}^{'}\mathbf{Y}^{'}(\mathbf{P}_Z - \mathbf{P}_{Z_1})\mathbf{Y}\boldsymbol{\beta}}{\boldsymbol{\beta}^{'}\mathbf{Y}^{'}\bar{\mathbf{P}}_Z\mathbf{Y}\boldsymbol{\beta}} \tag{28}$$

を最小化する解 (最小分散比推定量) が考えられる。この解は $(1+G_2)\times 1$) ベクトル $\boldsymbol{\beta}$ の二次形式の比となることから (27) の最小固有値に対応する $l_n(n/q_n)$ に一致することがわかる。この解が階数条件の下で最尤推定になっていることについては、例えば国友 (2011)4.2 節の証明が比較的簡単である。なお Anderson (1951) はここでの問題から reduced rank regression(縮

小階数回帰) の理論を開発している。
構造方程式の係数ベクトル $\boldsymbol{\beta}$ の二段階最小二乗 (TSLS) 推定量 $\hat{\boldsymbol{\beta}}_{TS}$ $(= (1, -\hat{\boldsymbol{\beta}}_{2.TS}^{'})')$ は

$$\mathbf{Y}_2'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{Y}\begin{pmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{2.TS} \end{pmatrix} = \mathbf{0}\ , \tag{29}$$

で与えられるが、$\mathbf{Y}_2 = (\mathbf{y}_i')$ は $n \times G_2$ 行列である。この TSLS 推定量は分散比の分子を最小化する解に一致する。LIML は最小分散比推定に一致して

$$\left[\mathbf{Y}_2'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{Y} - (\frac{n}{q_n})\lambda_1\mathbf{Y}_2'(\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{Y}\right]\begin{pmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{2.LIML} \end{pmatrix} = \mathbf{0}\ , \tag{30}$$

と表現される。なお構造方程式の係数ベクトル $\boldsymbol{\gamma}_1$ の LIML 推定量と TSLS 推定量は

$$\hat{\boldsymbol{\gamma}}_1 = (\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathbf{Y}\hat{\boldsymbol{\beta}}\ , \tag{31}$$

で与えられるが、ここで $\hat{\boldsymbol{\beta}}$ は $\hat{\boldsymbol{\beta}}_{LI}$ あるいは $\hat{\boldsymbol{\beta}}_{TS}$ である。

操作変数ベクトル $(G_2 \times 1)$ を $\mathbf{w}_{2i}$ とする。$n \times (K_1 + G_2)$ 行列 $\mathbf{W} = (\mathbf{Z}_1, \mathbf{W}_2) = (\mathbf{z}_{1i}', \mathbf{w}_{2i}')$ （行列の階数は退化せず $K_1 + G_2$）とすると、(i) 条件 $\frac{1}{n}\mathbf{W}'\mathbf{Z} \xrightarrow{p} (\mathbf{M}_{z_1,z}, \mathbf{M}_{w_2,z})'$ (確率 1 で階数は $K_1 + G_2$, すなわちランク落ちしない)、(ii) $\frac{1}{n}\mathbf{W}'\mathbf{V} \xrightarrow{p} \mathbf{O}$ (ゼロ行列に確率収束する) と云う条件を仮定する必要がある。
TSLS, LIML における対応する条件は $\mathbf{W}$ を $\mathbf{Z}$ として $(i)'$ 条件 $\frac{1}{n}\mathbf{Z}'\mathbf{Z} \xrightarrow{p} \mathbf{M}_{z,z}$ (確率 1 で階数は K_1+G_2, すなわちランク落ちしない)、$(ii)'$ $\frac{1}{n}\mathbf{Z}'\mathbf{V} \xrightarrow{p} \mathbf{O}$ (ゼロ行列に確率収束する)、である。
操作変数推定量 (Instrumental Variables Estimator, 略して IV) は

$$\begin{pmatrix} \mathbf{Z}_1' \\ \mathbf{W}_2' \end{pmatrix}(\mathbf{Z}_1, \mathbf{Y})\begin{pmatrix} -\boldsymbol{\gamma}_1 \\ 1 \\ -\hat{\boldsymbol{\beta}}_{2.IV} \end{pmatrix} = \mathbf{0}\ , \tag{32}$$

で表現される。なお、LIML 推定量や TSLS 推定量は操作変数の表現を持ち、条件を満たすので LIML, TSLS はともに操作変数と解釈が可能であるが、操作変数法の中では漸近的には効率的になる[16]。

[16]補論の命題 3 を参照されたい。なお古典的な証明では分散均一性を仮定しているがより一般的な場合でも成立すると思われる。

構造方程式の母係数の推定量 $(\hat{\boldsymbol{\beta}}_2', \hat{\boldsymbol{\gamma}}_1')'$ $((G_2+K_1)\times 1$ ベクトル) の漸近分布は Anderson-Rubin(1949, 1950) により導かれた。ここで

$$\mathbf{B}_n = \left[\begin{array}{c} \mathbf{\Pi}_2' \\ (\mathbf{I}_{K_1}, \mathbf{O}) \end{array} \right] \frac{1}{n}\mathbf{Z}'\mathbf{Z} \left[\mathbf{\Pi}_2, \left(\begin{array}{c} \mathbf{I}_{K_1} \\ \mathbf{O} \end{array} \right) \right] ,$$

$$\mathbf{C}_n = \left[\begin{array}{c} \mathbf{\Pi}_2' \\ (\mathbf{I}_{K_1}, \mathbf{O}) \end{array} \right] \frac{1}{n}\mathbf{Z}'\hat{\mathbf{u}}\hat{\mathbf{u}}'\mathbf{Z} \left[\mathbf{\Pi}_2, \left(\begin{array}{c} \mathbf{I}_{K_1} \\ \mathbf{O} \end{array} \right) \right] ,$$

とする。ただし $\hat{\mathbf{u}}$ は残差ベクトル $\hat{\mathbf{u}} = (\mathbf{Z}_1, \mathbf{Y})(-\hat{\boldsymbol{\gamma}}_1', 1, -\hat{\boldsymbol{\beta}}_{2.IV}')'$ というように母係数の推定量を用いて構成し、さらに

$$\mathrm{plim}_{n\to\infty}\mathbf{B}_n = \mathbf{B} > 0 \text{ (positive definite)} ,\ \mathrm{plim}_{n\to\infty}\mathbf{C}_n = \mathbf{C} \ ,$$

が存在するとしよう。標準的な仮定の下での推定量の漸近的性質については次の結果が知られている[17]。

命題 2 : 構造方程式モデル (14), (15) において非退化条件 $(1/n)\mathbf{Z}'\mathbf{Z} \xrightarrow{p} \mathbf{M}$（正定符号）、識別条件 (17) ($\mathbf{\Pi}$ の階数は G_2) を仮定する。また誤差項 $\mathbf{v}_i((1+G_2)\times 1$ ベクトル) は異なる i $(1,\cdots,n)$ について互いに独立で $\mathbf{E}[\mathbf{v}_i|\mathbf{z}_i] = \mathbf{0}$, 任意の $\delta > 0$ に対し $\mathbf{E}[\|\mathbf{v}_i\|^{1+\delta}] < +\infty$ となる (分散共分散は存在する)、などの正則条件を満たすものとする。このとき TSLS 推定量と LIML 推定量はともに漸近的に次の正規分布に収束する。

$$\sqrt{n} \left[\begin{array}{c} \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1 \end{array} \right] \xrightarrow{w} N[\mathbf{0}, \mathbf{B}^{-1}\mathbf{C}\mathbf{B}^{-1}] . \tag{33}$$

二つの推定量 TSLS, LIML は通常の漸近理論の仮定の下では $\lambda_1 \xrightarrow{p} 0$ より漸近的には同等となり、ある種の漸近限界に一致する。なお分散均一の場合には $\mathbf{E}[\mathbf{u}\mathbf{u}'] = \sigma^2\mathbf{I}_n$ より漸近分散共分散は $\sigma^2\mathbf{B}^{-1}$ で与えられる。また、K_2 が大きいと TSLS に比べて LIML の優位性が高まることが知られている。

なお分散共分散行列に現れる行列 $\mathbf{B}_n$ と $\mathbf{C}_n$ は行列 $\mathbf{\Pi}_2$ の最小二乗推定値 ($\mathbf{y}_{2i}$ を $\mathbf{z}_i$ に回帰する) および構造方程式の残差推定値 $\hat{\mathbf{u}}$ より構成すれ

[17]古典的な文献では分散・共分散の均一性を仮定しているが、より一般的な場合にも成立する。

ば、通常の仮定の下で一致性がある[18]。

漸近的ロバストネス：
ここで述べた漸近的結果はパラメター数は固定した下でデータ数 n が大きくした場合であり、このとき誤差項についての元の分布形の仮定に関わらず漸近的結果は成立することに特に注意しておく。例えば $G_2 = 1$、2 値をとる内生的ダミー変数の場合には y_{2i} は 0,1 の値のみをとるので対応する誤差項 v_{2i} の分布は離散分布にしたがう分布は離散分布であるが、AIR (1996) が考察している場合でも本節の主張は成立する。近年の疫学・臨床試験でも n が大きい場合が少なくないので重要な意味があると思われる。例えば LIML は元々は誤差が正規分布にしたがう場合に導出されたが、こうした漸近的ロバストネスは一般的な条件の下で成り立つことのである。

なお確率分布を特定化できれば AIC など尤度関数にもとづくモデル選択法を利用することが可能であるが、漸近的意味で正規分布に基づくモデル選択も意味があると考えられる。誤差項に関する正規分布の仮定に基づく対数尤度関数は

$$\log L \propto -\frac{n}{2}\log[(1+\lambda_1)|\mathbf{H}|] \tag{34}$$

に比例する[19]。データが極端に少なくなければ誤差の正規性がなくとも AIC 最小化によるモデル選択規準を構成するなど、計量生物ではあまり知られていないようであるが、有用な情報ではないかと予想される。

検定問題：
さらに構造方程式の推定問題の開発とともに幾つかの統計的仮説検定、モデル選択も問題も重要であり係数の有意性、過剰識別性 (over-identification)、先決性（predeterminedness, あるいは外生性 exnogeneity）などについて仮説検定問題が議論されていた。この中で (14) 式で表される構造方程式の母係数についての Anderson-Rubin 統計量は Anderson-Rubin (1950) が開発している。また (17) 式で表されている過剰識別性条件についてお検定問題については例えば Anderson-Kunitomo (1992) など多くの文献で議論されている。

---

[18] ここでは詳しく述べないが、識別条件下の $\mathbf{\Pi}_2$ は制約があるが詳しくは森棟 (1985) を参照されたい。

[19] 構造方程式の推定を巡る古典的な統計的展開については森棟 (1985) を参照されたい。

5　同時方程式と操作変数法

操作変数法 (Instrumental Variables Method, IV 法) が統計的因果推論に登場するようになり、統計家の関心が高まるにつれ、例えば Imbens (2014) のような計量経済関係者による解説記事が登場するようになった。しかしながら AIR (1996) や Imbens (2014) の解説はより旧世代の関係者から見ると一面的でありかなり偏っていると考えられるので、なぜそう見るのか幾つかの論点を説明することにする。

同時方程式や構造方程式の統計モデルが今でも経済学において市場経済の基本的な分析手段である需要曲線と供給曲線を観測データから推定しようとした時に生じる問題から発生したことは間違いない。この場合、通常の線形回帰問題と異なる識別問題を指摘したのは Working (1927) であると思われる[20]。。次に登場するのは Wald (1940), Durbin (1954) であるが、これにはかなりの飛躍があり、構造方程式の統計理論と応用の展開をかなり無視していると考えられる。Wald (1940) は変数誤差 (errors-in-variables) 問題を解決する統計的方法として Wald 法を開発した。しかし、より一般的に数理統計学の問題として定式化したのは Havelmo (1944) の研究に触発され当時の Cowles Commission（シカゴ大学コールズ研究所）に所属していた Anderson-Rubin (1949, 1950) であり、その辺の事情は Anderson (2015) に詳しい。彼らは同時方程式の推定理論として単一方程式（つまり構造方程式）の最尤法を考察し、制限情報最尤法 (limited information maximum likelihood, LIML) を開発、統計的推測理論の一般論を統計的多変量解析の問題として展開したが、これは当時のマクロ経済モデルの開発という経済政策上の要請に答えるという側面もあった。LIML 法では固有値問題を数値的に解く必要があり、当時の経済学における計算環境の貧弱性もあり、簡便な統計的方法の開発が望まれていた。この問題について Durbin (1954) は操作変数法 (insturumental variables method, IV 法) を構造方程式とは同値の変数誤差問題を解決する手段として開発、Theil (1954, unpublished) は二段階最小二乗法 (two-stage least squares method, TSLS) を開発した。その後、特にデータ数 n が大きくない場合や除かれた外生変数 K_2 が大きい場合などについて、OLS, TSLS, LIML の有限標本の性質などを巡る統計的議論がかなり盛んに行われるようになり、T.W. Anderson, 竹内啓 (Takeuchi, K.), 佐和隆光 (Sawa,T.), 森棟公

[20] AIR (1996) や Imbens (2014) は Wright (1928) に始まると説明している。この辺の事情や説明は米国ボストン周辺の econometricians の理解と一昔前のスタンフォードなどの西海岸の econometricians の理解は微妙に異なると思われる。

夫 (Morimune, K.), 国友直人 (Kunitomo, N.) などがかなりの数の研究報告を行ったが、例えば Anderson-Sawa (1979), Anderson-Kunitomo-Sawa (1982), Anderson-Kunitomo-Morimune (1986) などがある。

2022 年の時点で振り返っても幾つかの統計的問題を考察するにあたって、Heckman などによる別の定式化など様々な研究が行われたが、統計的因果推論では重要な内生変数がダミー変数の場合などについてはなお十分には理解されていないと思われる。非線形構造方程式についての古典的議論は Amemiya (1985) にまとめられているが、より最近の幾つかの研究については Imbens (2010) が言及している。現時点では計量生物を含め、応用上で有用な結果が得られたとの理解はないようである。

6 部分識別問題

AIR (1996) が一つの契機となり計量経済分野では統計的因果推論を巡り多くの理論的・応用的な研究が行われるようになっている。中でも識別性を巡る問題に基本問題に一石を投じたのが Manski の部分識別 (partial identification) を巡る一連の研究である。古典的な計量経済では構造方程式の母数が識別可能 (identifiable) となる条件、推定法の開発に力を注いでいた。Manski によれば一般には識別性をもたらす仮定を無理に置くよりも、識別性に依存することなくデータに矛盾しないような信頼集合[21] (credible region, set) を導くことの重要性を指摘しているのである。例えば Rubin モデルにおいてランダムに割付け（$D_i = 1, 0$）が可能でないとき、観測データから ATE を推定するには $\mathbf{E}[Y_i(1)], \mathbf{E}[Y_i(0)]$ を推定する必要がある。このとき表現

$$\mathbf{E}[Y_i(1)] = \mathbf{E}[Y_i(1)|D_i = 1]P(D_i = 1) + \mathbf{E}[Y_i(1)|D_i = 0]P(D_i = 0),$$

$$\mathbf{E}[Y_i(0)] = \mathbf{E}[Y_i(0)|D_i = 1]P(D_i = 1) + \mathbf{E}[Y_i(0)|D_i = 0]P(D_i = 0)$$

において、$\mathbf{E}[Y_i(1)|D_i = 0]$,$\mathbf{E}[Y_i(0)|D_i = 1]$ は観測されない。特に Y_i の値域が有界で $\underline{Y} \leq Y \leq \bar{Y}$ の場合には、観測から識別できない項を $\underline{Y}$,$\bar{Y}$ で置き換えると

$$ATE_L \leq \mathbf{E}[Y_i(1)] - \mathbf{E}[Y_i(0)] \leq ATE_U \tag{35}$$

[21]帰無仮説はある種の複合仮説と理解すると、古典的統計学における信頼区間 (confidence set) やベイズ統計学における信用区間 (credible set) など様々な可能性が考えられる。

という境界 (bound) が得られる。

Manski の主な主張は点識別 (point identification) が正確に成り立たない場合に統計的モデルを利用すると、しばしば同じ対象についての実証分析が異なることになることなどへの一つの説明の可能性などという科学的な意味においては一考すべきではある。しかしながら本筆者は今のところ、一般的な場合に応用上で意味がある境界が導けるかはかなり懐疑的である[22]。目的変数が有界でないと意味を持たない上に、Manski (1997), Manski and Pepper (2000) では Monotone Treatment Response (MTR), Concave MTR などの条件による Sharp Bound を構成しようとしているが、結局は関数についてのある種の条件を仮定していることになっている。例えば奥村 (2018) が説明しているような賃金データを利用して高卒、大卒、院卒という学歴の所得に対する影響は平均的には必ず増加関数と先験的に仮定することに妥当性があるか、十分に説得的である、とは言い難い。しかしながら、従来の点識別に対して区間識別、部分集合識別 (set identification）下での推定問題というテーマの重要性は今後さらに注目すべきものとなろう[23]。

7　再び因果推論と構造方程式を巡って

計量経済 (econometrics) 分野はその出発時点から市場経済で観測されるデータ分析では (Working (1927) が直観的に指摘したように)、前世紀から統計理論家が開発した相関や回帰といった統計的方法をそのまま利用すると問題が生じることに気が付いた。需要関数、供給関数といった構造方程式の推定問題は Haavelmo (1944) の研究を一つの重要な契機として研究が発展した。例えばあるエコノミストによる経済データの統計的分析に対して「(経済) 理論なき計測 (measuremnt without theory)」(Koopmans (1947)) などのコメントが浴びせたりしたが、そうした議論の中で操作変数法をはじめとする統計的方法が開発された。しかし、こうした計量経済における構造方程式の研究に対する当時の統計家の典型的反応は

[22] 例えば ATE 推定において Noncompliance が存在しない場合は信頼区間を構成でき、データ数が大きければ区間幅は短くなる。もっとも現実の例としては、内閣府当局者は GDP の伸び率について、かなり昔から昔から低信頼度が指摘されているにも関わらず、点推定値のみを報告、信頼区間は発表していない。それほど多くない時系列データから「景気が悪くなったか良くなったか正直分からない」のでは当局の信用に関わるのであろう。

[23] 関連する諸問題については例えば奥村 (2018) が説明している。

“*I despair of ever understanding the logic of simultaneous equations well enough to tackle them,*”
“*Additionally, some variables are taken to be exogenous (independent of the disturbance terms) and some endogenous (dependent on the disturbance terms). The rationale is seldom clear, because—among other things—there is seldom any very clear description of what the disturbance terms mean, or where they come from*”
などであった[24]。

こうした流れは AIR(1996) 以来、劇的に変化し、例えば臨床試験で普通に操作変数法が議論され、公共政策の評価では反実仮想モデルが標準になっている。今や統計的因果推論や操作変数法の応用では生物統計と計量経済はかなり類似の発想で議論する時代へと変貌を遂げている。生物統計において議論されるようになった反実仮想という考え方は実は計量経済における構造方程式の議論そのものなのである。また、計量経済分野では新たに micro-econometrics (ミクロ計量経済) 分野が発展を遂げ、現代の応用経済学では統計的因果推論の基本は既に実証分析では必須の内容になっている。

本稿でとりあげている問題設定は、一般的に構造方程式に現れる内生変数 $(\mathbf{y}_{1i}^{`},\mathbf{y}_{2i}^{`})^{`}$ $((G_1+G_2)\times 1$ ベクトル) と操作変数 (あるいは外生変数) $\mathbf{z}_i$ ($K\times 1$ ベクトル)$(i=1,\cdots,n)$ の同時分布が利用可能であり $G_1=1$ の場合である。$G_1\geq 1$ の場合への拡張は可能であり構造方程式の部分システム法と呼ばれている[25]。この同時分布から実ベクトル $\mathbf{w}_j$ ($K_2\times 1$ ベクトル, $K_2\leq K$) $(j=1,\cdots,m)$ に対して操作変数（あるいは外生変数）を所与とする $\mathbf{y}_{1i}$ の条件付分布

$$\mathrm{P}(\mathbf{y}_{1i}\leq \mathbf{y}_1|\mathbf{z}_{2i}=\mathbf{w}_j)=\int_{y_{2i}}\mathrm{P}(\mathbf{y}_{1i}\leq \mathbf{y}_1,\ \mathbf{y}_{2i}|\mathbf{z}_{2i}=\mathbf{w}_j)dF(\mathbf{y}_{2i}|\mathbf{z}_{2i}=\mathbf{w}_j) \tag{36}$$

$(i=1,\cdots,n:j=0.1,\cdots,m)$ を比較する統計的問題と見なすことができる。Imbns-Rubin(2015) で説明している統計的因果推論の表現と本稿で用いた構造方程式の表現や用語は同一ではないが、本稿での記号を利用して $m=1$,$G_1=G_2=K_1=K_2=1$ とすると、AIR(1996) の Noncompliance

[24]Imbens (2014) から例を二つ引用した。後者の David Freedman のコメントは直接にセミナーで聞いたことがある。

[25](26) に対応する母集団でのゼロ固有値の数が G_1 となるが 4 節の議論はほぼそのまま成立する。例えば Anderson-Kunitomo (1992) を参照されたい。

問題は $\mathrm{P}(y_{1i} \le y_1 | \mathbf{z}_{2i} = \mathbf{w}_1) - \mathrm{P}(y_{1i} \le y_1 | \mathbf{z}_{2i} = \mathbf{w}_0)$ を比較する問題と解釈可能である。したがつて、本稿での議論を踏まえると様々な方向での利用が可能と考えられる。

他方、統計的因果推論の問題については様々な議論が考えられ、既に議論を紹介した AIR(1996) の Noncompliance 問題や Manski の主張とは異なる方向として、例えば計量経済学では Heckman (2008) は社会における政策評価分野では異なる主体に対する政策効果の予測が重要であり、Rubin 流の反実仮想モデルは不十分であるであることを主張している。例えば Heckman (2008), Heckman and Vytlacil (2005), Heckman, J. Urzua and E. Vytlacil (2006), は Marginal Treatment Effects(MTE, 限界処置効果) と云う独自の因果効果量と政策評価法を提案していることにも言及する必要があろう。

8 展望

Rubin (1974) による統計的因果モデルは疫学での臨床試験や経済学・公共政策での政策評価に対する応用などにおいて、それまでの実験データの解析を念頭にした統計学の発想に加え、新たな統計的アプローチの門を開いた。反実仮想 (counter factual) モデルは単純な統計モデルであるが因果推論という難しい問題に対して統計学の観点から新たな光をあて、AIR(1996) による操作変数の利用は統計科学における生物統計と計量経済の壁を取り除く試みの第一歩と評価される。

他方、疫学や政策評価などの応用の観点からは十分な道具立てとは云えず、幾つかの重要な方向に拡張する必要がある。AIR(1996) の Noncompliance モデルは構造方程式モデルにおいて 1 個の除かれた外生変数が割付け変数、ダミー変数の場合と解釈できるが、Noncompliance モデルは様々な方向に拡張可能である。例えば、ノンコンプライアンス問題、観測データには計測誤差が存在するのは一般的な状況であり、性別や年齢と云った説明変数、2 水準以上の内生的ダミー変数、複数の操作変数などを考慮する必要がある。こうした問題を解決する手段としては操作変数法を含む構造方程式の統計モデルが有効であると考えられる。

ここで、より一般的な統計モデルを視野に入れると、処理効果の識別問題の扱いをはじめとする幾つかの重要な未解決な統計的課題があることが浮上してくる。社会・経済などにおける政策評価などではランダム化比較実験の役割はかなり限定的であり、厳密な意味でそれを実施する

ことは困難である場合にも適用可能という意味では構造方程式モデルは有用と考えられる。この論点に関連して近年では例えば公共政策についての Imbens (2010)、統計的因果推論についての Imbens (2014) のサーベイで議論しているが、パネルデータ分析などは一つの有力なアプローチだろう。

さらに、最近では大量のデータ (big data) による個人レベルの処置効果の不均一性、異質的処置効果 (Heterogeneous Treatment Effects, HTE) を巡る統計的分析が盛んである。中でも Random Forests という統計的学習論 (statistical learning, machine learning) にもとづく Wager-Athey (2018), Athey-Tibshirani-Wager (2019) の Causal Forests、Abbadie (2021) による Synthetic Control 政策評価法、などが注目される。

最後になるが計量生物、計量経済 (あるいは公共政策)、などの分野では現実に利用可能なデータを用いた統計分析にはなお様々な課題がある。例えば実際の臨床試験や政策評価は 1 度の実験で処置効果を計測するばかりではなく、複数回行うことがかなり一般的であるから (3 節の需給均衡の例でほんの少し議論したように) 時間効果や複数の計測結果 ($G_1 > 1$) を同時に考慮する場合の臨床試験の設計や統計的因果分析なども重要であろう。むろん本稿はこれまで見過ごされたと筆者が考えた構造方程式と操作変数の利用を巡る幾つかの論点を題材に一石を投じたにすぎない。本稿で議論したように、統計科学における計量生物と計量経済での発展を意識しつつ、統計的因果推論と構造方程式を巡る操作変数の利用した統計的分析は今後にさらなる展開が期待される。

参考文献

[1] Abadie, A. (2021), "Using synthetic controls: Feasibility, data requairements, and methodological aspects," *Journal of Economic Literature*, 59 , 391425.

[2] Amemiya, T. (1985), *Advanced Econometrics*, Blackwell.

[3] Anderson, T.W. (2003), *An Introduction to Multivariate Statistical Analysis*, 3rd Edition, Wiley.

[4] Anderson, T.W. (2015), "My reminiscences of Trygve Haavelmo at

the Cowles Commission," *Econometric Theory*, 180-192.

[5] Anderson, T.W. and H. Rubin (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stocchastic Equations," *Annals of Mathematical Statistics*, 46-63.

[6] Anderson, T.W. and H. Rubin (1950), "The Asymptotic Properties of Estimates of the the Parameters of a Single Equation in a Complete System of Stocchastic Equations," *Annals of Mathematical Statistics*, 570-582.

[7] Anderson, T. and T. Sawa (1979), "Evaluation of the Distribution Function of the Two-stage Least Squares Estimate," *Economerica*, Vol. 47, 163-182.

[8] Anderson, T.W., N. Kunitomo and T. Sawa (1982), "Evaluation of the Distribution Function of the Limited Information Maximum LLikelihood Estimator," *Economerica*, Vol. 50, 1009-1026.

[9] Anderson, T.W., N. Kunitomo and K. Morimune (1986), "Comparing Single-Equation Estimators in a Simultaneous Equation System," *Economeric Theory*, Vol. 2, 1-32.

[10] Anderson, T.W. and N. Kunitomo (1992), "Tests of overidentification and predeterminedness in simultaneous equation models," *Journal of Econometrics*, Vol. 54, 49-78.

[11] Anderson, T.W., N. Kunitomo and Y. Matsushita (2010), "On the asymptotic optimality of the LIML estimator with possibly many instruments," *Journal of Econometrics*, Vol.157, 191-204.

[12] Angrist, J., Imbens, G. and Rubin, D (1996), "Identification of causal effects using instrumental variables (with discussions)," *Journal of the American Statistical Association*, Vol. 91, 444-472.

[13] Athey, S., J. Tibshirani and Wager (2019), "Generalized Random Forests," *Annals of Statistics*, Vol. 47-2, 1148-1178.

[14] Durbin, J. (1954), "Errors In Variables," *Review of the International Statistics*, Vol. 22, 23-32.

[15] Haarvelmo, T. (1944), "The Probability Approach to Econometrics," *Econometrica*, Vol. 12, 1-115.

[16] Hayashi, F. (2000), *Econometrics*, Princeton University Press.

[17] Heckman and Vytlacil (2005), "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, Vol. 73-3, 669-738.

[18] Heckman, J. Urzua and E. Vytlacil (2006), "Understanding Instrumental Variables in Models with Essential Heterogeneity," *The Review of Economics and Statistics*, Vol. 88-3, 389-432.

[19] Heckman, J. (2008), "Econometric Causality," *NBER Working paper Series 13934*, ISI Conference.

[20] Imbens, G. (2010),"Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Perspectives*, Vol. 47-1, 5-86.

[21] Imbens, G. (2014), "Instrumental Variables," *Statistical Science*, Vol. 29, 323-379, with Comments and Rejoinder.

[22] Imbens, G. and J. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, Vol. 62, 467-475.

[23] Imbens, G. and D. Rubin (2015), *Causal Inference for Statistics, Social and Biomedical Sciences*, Cambridge University Press.

[24] Kitagawa, T. (2015), "A Test for Instrumental Validity," *Econometrica*, Vol. 83-5, 2043-2063.

[25] Koopmans, T. (1947), "Measurement Without Theory," *The Review of Economics and Statistics*, Vol. 29-3, 161-172.

[26] Kunitomo, N. and S. Sato (1996), "Asymmetry in economic time series and simultaneous switching autoregressive models," *Structural Change and Economic Dynamics*, Vol. 7, 1-34.

[27] Manski, C. (1997), "Monotone Treatment Response," *Econometrica*, Vol. 65-6, 1311-1334.

[28] Manski, C. and J. Pepper (2000), "Monotone Instrumental Variables with an application to the Returns to Schooling," *Econometrica*, Vol. 68-4, 997-1010.

[29] Quandt, R. E. (1988), "Econometrics of Disequilibrium," Blackwell.

[30] Theil, H. (1971), *Principle of Econometrics*, Wiley.

[31] Rubin, D. (1974), "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, Vol. 66, 688-701.

[32] Sargan, D. (1958), "The estimation of economic relathionships using instrumental variales," *Econometrica*, Vol. 26, 393-415.

[33] Wager, S. and S. Athey (2018), "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests," *Journal of the American Statistical Association*, Vol. 113, 41, 1228-1242.

[34] Wald, A. (1940), "The Fitting Straight Lines if Both Varaibles are subject to Error," *Anaals of Mathematical Statistics*, 284-300.

[35] Working, E. (1927), "What Statistical Demand Curve Shows ?," *Quarterly Journal of Economics*, Vol. 41, 212-235.

[36] Wright, S. (1928), "Appendix in The Tariff on Animal and Vegitable Oils," MacMillan.

[37] 岩崎学 (2015), 統計的因果推論, 朝倉書店。

[38] 奥村綱雄 (2018), 部分識別入門, 日本評論社。

[38] 国友直人 (2011)「構造方程式モデルと計量経済学」, 朝倉書店。

[40] 森棟公夫 (1985), 経済モデルの推定と検定, 共立出版。

[41] 西山慶彦・新谷元嗣・川口大司・奥井亮 (2019), 計量経済学, 有斐閣。

補論 A

この補論では 2 節における主張を示す。導出はいずれも初等的な議論によるが、念のために説明しておく。

(i) (2) 式を変形すると

$$\hat{\beta} = \frac{\sum_{i=1}^n Z_i Y_i \sum_{i=1}^n (1-Z_i) - \sum_{i=1}^n (1-Z_i) Y_i \sum_{i=1}^n Z_i}{\sum_{i=1}^n Z_i \sum_{i=1}^n (1-Z_i)} \tag{A.37}$$

である。ここで Z_i はダミー変数であるから ($Z_i^2 = Z_i$)、分母は $n[\sum_{i=1}^n Z_i^2 - n\bar{Z}^2]$ ($\bar{Z} = (1/n)\sum_{i=1}^n Z_i$), 分子は $n[\sum_{i=1}^n Z_iY_i - n\bar{Z}\bar{Y}]$ ($\bar{Y} = (1/n)\sum_{i=1}^n Y_i$), となるので $\hat{\beta} = \hat{\beta}_{LS}$ である。

(ii) 2 節の主張 $d = 1$ or 0 のとき $\mathbf{E}[u_i|Z_i = d] = 0$ であっても、一般には $\mathbf{E}[u_i|D_i = d] \neq 0$ となることを説明する。D_i の方程式を係数 π_{12}, π_{22} を用いて $D_i = \pi_{12} + \pi_{22}Z_i + v_{2i}$ と表そう。ここで v_{2i} は $\mathbf{E}[v_{2i}|\mathbf{z}_i] = 0$ となる誤差項（確率変数）である。この式を $Y_i = \alpha + \beta D_i + u_i$ $(i = 1, \cdots, n)$ に代入すると

$$Y_i = \pi_{11} + \pi_{21}Z_i + v_{1i} \ , \tag{A.38}$$

ただし $\pi_{11} = \alpha + \beta\pi_{12}$, $\pi_{21} = \beta\pi_{22}$, $v_{1i} = u_i + \beta v_{2i}$ である。
ここで $p_{lk} = P(Z_i = l, D_i = l)$, $(k, l = 0, 1)$ とすると条件付期待値の条件 $\mathbf{E}[v_{2i}|Z_i = 1] = p_{11}(1 - \pi_{12} - \pi_{22}) + p_{10}(-\pi_{12}) = 0$, $\mathbf{E}[v_{i2}|Z_i = 0] = p_{01}(1 - \pi_{12}) + p_{10}(-\pi_{12}) = 0$ を解くと $p_{01} + p_{00} = 1$ より $p_{01} = \pi_{12}$, $p_{11}(1 - \pi_{22}) - \pi_{12} = 0$ となる。
ここで事象 $D_i = 1$ を条件とする期待値を評価すると、$\mathbf{E}[v_{2i}|D_i = 1] = (1 - \pi_{12} - \pi_{22})p_{11} + (1 - \pi_{12})p_{01}$, $\mathbf{E}[v_{1i}|D_i = 1] = (\mathbf{E}[Y_i|D_i = 1] - \pi_{11} - \pi_{12})p_{11} + (\mathbf{E}[Y_i|D_i = 1] - \pi_{11})p_{01}$ となる。
この関係を利用すると（単純だが若干の変形が評価が必要)、$\mathbf{E}[u_i|D_i = 1] = \mathbf{E}[v_{1i}|D_i = 1] - \beta\mathbf{E}[v_{2i}|D_i = 1] = (\pi_{11} + \mathbf{E}[u_i|D_i = 1])(p_{10} + p_{00})$ となるので、結局

$$\mathbf{E}[u_i|D_i = 1] = \frac{\pi_{11}}{1 - (p_{10} + p_{00})} \tag{A.39}$$

が得られる。ここで $\pi_{11} \neq 0$ かつ $0 < P(D_i = 1) < 1$ ならば (A.39) はゼロとはならないことが分かる。

(iii) ここで (ii) を利用して β^* を評価してみよう。$\sigma_z^2 = \mathrm{Var}[Z_i]$, $\sigma_{2v}^2 = \mathbf{E}[v_{2i}^2]$ (ただし $\mathbf{E}[v_i|\mathbf{z}_i] = 0$) とすると $D_i = \pi_{12} + \pi_{22}Z_i + v_{2i}$ および (A.38) より

$$\beta^* = \frac{\mathrm{Cov}(Y_i, D_i)}{\mathrm{Var}(D_i)} = \frac{\beta\pi_{22}^2\sigma_z^2 + \beta\sigma_{2v}^2}{\pi_{22}^2\sigma_z^2 + \sigma_{2v}^2}$$

となる。観測誤差がなければ $\sigma_{2v} = 0$ より $\beta^* = \beta$ となるが、より一般には観測誤差があると $\sigma_{2v} > 0$, $\sigma_u^2 > 0$, $\pi_{22} \neq 0$ であるから両者は一致しない。

補論B

この補論Bでは操作変数推定量, TSLS, LIML, およびGMMの漸近的性質を議論する。(構造方程式の推定では最小二乗推定(OLS)は一致性がないのでここでは議論を省略する。)
以下ではnが大きい標準的な場合（分散均一な場合 etc.）についての命題3を示すが、より一般の場合も成立する。まず誤差項はi.i.d.系列で分散・共分散行列が均一の場合、TSLS推定量とLIML推定量は漸近的に効率的となることを主張する命題3とその証明の概略を説明する。

命題 3： 操作変数の非退化条件、識別条件(17) ($\mathbf{\Pi}$の階数はG_2)を仮定する($K_2 \geq G_2$を意味する)。また$\mathbf{z}_{1i}$に加えてG_2個の退化しない操作変数$\mathbf{W}_2(n \times G_2$を用意して(32)式より操作変数推定量を構成する。(i) $\frac{1}{n}\mathbf{W}'\mathbf{Z} \xrightarrow{p} (\mathbf{M}_{z_1,z}, \mathbf{M}_{w_2,z})'$ (確率1で階数はK_1+G_2)、(ii) 誤差項$\mathbf{v}_i((1+G_2)\times 1$ベクトル)はi.i.d.系列で$\mathbf{E}[\mathbf{v}_i|\mathbf{z}_i]=\mathbf{0}$, 分散共分散が有界、などの正則条件を仮定する。このとき操作変数推定量$\hat{\boldsymbol{\beta}}_{2.IV}$は漸近正規性を持ち、

$$\sqrt{n}\left[\hat{\boldsymbol{\beta}}_{2,IV} - \boldsymbol{\beta}_2\right] \xrightarrow{w} N[\mathbf{0}, \sigma^2\mathbf{D}] \ , \tag{A.40}$$

ただし

$$\mathbf{D} = \mathrm{plim}_{n\to\infty}(\frac{1}{n}\mathbf{W}_2'\mathbf{Z}_{2.1}\mathbf{\Pi}_{22})^{-1}(\frac{1}{n}\mathbf{W}_2'\bar{\mathbf{P}}_{Z_1}\mathbf{W}_2)(\mathbf{\Pi}_{22}'\frac{1}{n}\mathbf{Z}_{2.1}'\mathbf{W}_2)^{-1}$$

である。ただし射影行列$\mathbf{P}_{Z_1} = \mathbf{Z}_1(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1$, $\bar{\mathbf{P}}_{Z_1} = \mathbf{I}_n - \mathbf{P}_{Z_1}$を用いて$\mathbf{Z}_{2.1} = \bar{\mathbf{P}}_{Z_1}\mathbf{Z}_2$である。
このとき行列不等式（非負定符号）の意味で

$$\mathbf{D} \geq (\mathbf{\Pi}_{22}'\mathrm{plim}_{n\to\infty}\frac{1}{n}\mathbf{A}_{22.1}\mathbf{\Pi}_{22})^{-1} \tag{A.41}$$

が成立する。
ここで$(\mathbf{\Pi}_{22}'\mathrm{plim}_{n\to\infty}\frac{1}{n}\mathbf{A}_{22.1}\mathbf{\Pi}_{22})^{-1}$は(33)式の$\mathbf{B}^{-1}$の左上$G_2\times G_2$部分行列に対応する。なお、上の表現に現れる確率収束先は正則行列でランク(階数)落ちがないことを仮定する。

証明の概略：(i) 操作変数推定量の表現から

$$\mathbf{Z}_1'\mathbf{Z}_1\hat{\gamma}_1 - \mathbf{Z}_1'\mathbf{Y}\hat{\boldsymbol{\beta}} = \mathbf{0} \ , \mathbf{W}_2'\mathbf{Z}_1\hat{\gamma}_1 - \mathbf{W}_2'\mathbf{Y}\hat{\boldsymbol{\beta}} = \mathbf{0} \ ,$$

であるから、$\hat{\gamma}_1$ を消去すると

$$\mathbf{W}_2^{'}[\mathbf{Z}_1(\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}\mathbf{Z}_1 - \mathbf{I}_n]\mathbf{Y}\hat{\boldsymbol{\beta}} = \mathbf{0}\ ,$$

となる。$\mathbf{Y} = \mathbf{Z\Pi} + \mathbf{V}$ を代入，$\mathbf{P}_{Z_1} = \mathbf{Z}_1(\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}\mathbf{Z}_1$,
$\mathbf{Z}_{2.1} = (\mathbf{I}_n - \mathbf{P}_{Z_1})\mathbf{Z}_2\ (= \bar{\mathbf{P}}_{Z_1}\mathbf{Z}_2)$ となることを利用すると、

$$\mathbf{W}_2^{'}[\mathbf{Z}_{2.1}(\mathbf{\Pi}_{21}, \mathbf{\Pi}_{22}) + (\mathbf{I}_n - \mathbf{P}_{Z_1})\mathbf{V}]\begin{pmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_2 \end{pmatrix} = \mathbf{0}\ ,$$

したがって、行列の非退化性の条件から

$$\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \xrightarrow{p} \mathbf{0}$$

が得られる。次に表現

$$\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 = [\mathbf{W}_2^{'}\mathbf{Z}_{2.1}\mathbf{\Pi}_{22} + \mathbf{W}_2^{'}(\mathbf{I}_n - \mathbf{P}_{Z_1})\mathbf{V}_2]^{-1}[\mathbf{W}_2^{'}(\mathbf{I}_n - \mathbf{P}_{Z_1})\mathbf{V}\boldsymbol{\beta}]$$

を利用する。条件から

$$\frac{1}{n}\mathbf{W}_2^{'}(\mathbf{I}_n - \mathbf{P}_{Z_1})\mathbf{V}_2 \xrightarrow{p} \mathbf{O}$$

であるから

$$\sqrt{n}[\hat{\boldsymbol{\beta}}_2^{*} - \boldsymbol{\beta}_2] = [\frac{1}{n}\mathbf{W}_2^{'}\mathbf{Z}_{2.1}\mathbf{\Pi}_{22}]^{-1}[\frac{1}{\sqrt{n}}\mathbf{W}_2^{'}(\mathbf{I}_n - \mathbf{P}_{Z_1})\mathbf{u}]$$

の漸近分布を求めれば良い。また条件より $\frac{1}{n}\mathbf{W}_2^{'}\mathbf{Z}_{2.1}\mathbf{\Pi}_{22}$ が確率収束するので $\frac{1}{\sqrt{n}}\mathbf{W}_2^{'}(\mathbf{I}_n - \mathbf{P}_{Z_1})\mathbf{u}$ に CLT(中心極限定理) を用いると漸近分布が得られる。
(ii) 行列 $\mathbf{C} = \mathbf{Z}_{2.1}\mathbf{\Pi}_{22}$ とすると $\mathbf{I}_n - \mathbf{C}(\mathbf{C}^{'}\mathbf{C})^{-1}\mathbf{C}$ は射影行列で非負定符号行列となるので

$$(\mathbf{W}_2^{'}\mathbf{Z}_{2.1}\mathbf{\Pi}_{22})(\mathbf{\Pi}_{22}^{'}\mathbf{A}_{22.1}\mathbf{\Pi}_{22})^{-1}(\mathbf{\Pi}_{22}^{'}\mathbf{Z}_{2.1}^{'}\mathbf{W}_2) \leq \mathbf{W}_2^{'}\bar{\mathbf{P}}_{Z_1}\mathbf{W}_2$$

が成立する。条件より行列の正則性が成り立つので非負定符号の意味の不等式が得られる。
(Q.E.D.)

計量経済学では分散が不均一な場合 (hetetoscedastic case) などに対応するときに GMM (generalized method of moments) がしばしば用いられて

いる。あまり文献では議論されていないようなので、TSLS, LIML, GMM の関係を本稿の枠組みの中で述べておく。構造方程式の直交条件 $(K \times 1)$ $\mathbf{E}[u_i|\mathbf{z}_i] = 0 \ (i = 1, \cdots, n)$ から

$$g_n(\boldsymbol{\delta}) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_i [y_{1i} - \mathbf{w}_i^{'} \boldsymbol{\delta}] \ , \tag{A.42}$$

とする。ただし $\mathbf{w}_i^{'} = (\mathbf{z}_{1i}^{'}, \mathbf{y}_{2i}^{'})$, $\boldsymbol{\delta}^{'} = (\boldsymbol{\gamma}_1^{'}, \boldsymbol{\beta}_2^{'})$ とする。適当な行列 $\mathbf{C}$（正定符号）をとり、評価関数 $L_n(\boldsymbol{\delta}) = [g_n(\boldsymbol{\delta})]^{'} \mathbf{C} [g_n(\boldsymbol{\delta})]$ を最小化する推定法は GMM 法と呼ばれている。特に n が大きい時に漸近共分散を最小にするような方法である効率的 GMM 推定量 (efficient GMM) の漸近分散は

$$\mathrm{AV}(\hat{\boldsymbol{\delta}}_{EGMM}) = [\boldsymbol{\Sigma}_{wz} \mathbf{S}^{-1} \boldsymbol{\Sigma}_{wz}^{'}]^{-1} \tag{A.43}$$

で与えられる。
ただし $\boldsymbol{\Sigma}_{wz} = \mathrm{plim}_{n \to \infty} \frac{1}{n} \sum \mathbf{z}_i \mathbf{w}_i^{'}$, $\mathbf{S} = \mathrm{plim}_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \hat{u}_i^2 \mathbf{z}_i \mathbf{z}_i^{'}$, であり、階数はそれぞれ $K_1 + G_2$, K となり退化しないことを仮定する[26]。$\boldsymbol{\Sigma}_{wz}$ は $(K_1 + G_2) \times K$ なので非退化性は $\mathrm{rank}(\boldsymbol{\Sigma}_{wz}) = K_1 + G_2$、特に丁度識別できる (just-identified) 場合には $K = K_1 + G_2$、したがって $K_2 = G_2$ の場合に対応する。

母数 (パラメター) を固定、$G_2, K \ (= K_1 + K_2)$ を固定、n が大きい時には次の漸近的結果が得られる。

命題 4： 操作変数の非退化条件、識別条件 (17) ($\boldsymbol{\Pi}$ の階数は G_2) を仮定する ($K_2 \geq G_2$ を意味する)。また誤差項 $\mathbf{v}_i((1 + G_2) \times 1$ ベクトル) はマルチンゲール差分系列で $\mathbf{E}[\mathbf{v}_i|\mathbf{z}_i] = \mathbf{0}$, 分散共分散が存在、行列の階数について正則条件を仮定する。
(i) $n \longrightarrow \infty$ のとき GMM 推定量 $\hat{\boldsymbol{\delta}}_{GMM}$, TSLS 推定量 $\hat{\boldsymbol{\delta}}_{TSLS}$, LIML 推定量 $\hat{\boldsymbol{\delta}}_{LIML}$ はいずれも漸近正規性を持つ。
(ii) 特に効率的 GMM 推定量 $\hat{\boldsymbol{\delta}}_{EGMM}$ の漸近共分散行列 $\mathbf{A}_1$, TSLS 推定量と LIML 推定量 $\hat{\boldsymbol{\delta}}_{LIML}$ の漸近共分散行列 $\mathbf{A}_2$ とすると次の関係が成立する。
(iia) $G_2 = K_2$ であれば $\mathbf{A}_1 - \mathbf{A}_2 = \mathbf{O}$,
(iib) $G_2 \leq K_2$ であれば非負定符号の意味で $\mathbf{A}_1 \leq \mathbf{A}_2$ となる。

なお、上の命題に関連して次のことが知られている。

[26]GMM についての議論は例えば本稿と記号が若干異なるが Hayashi (2000) を参照されたい。

(iii) 誤差項の分散が一定（均一分散）なら EGMM は TSLS に一致する。
(iv) K_2 が n とともに大きくなる場合 (many instruments の場合) は TSLS 推定量と GMM 推定量の一致性は失われるが、LIML は一致性を持ち、さらに漸近正規性も成り立つ[27]。すなわち一般の場合には Anderson-Kunitomo-Sawa (1982) が既に有限標本について指摘したように、漸近的にはバイアスと分散のトレードオフが存在している。

$G_1 \geq 1$ **の場合**： ここで左辺の内生変数 $\mathbf{y}_{1i}$ が $G_1 \geq 1$ ベクトルの場合について簡単に言及しておく。(14) は G_1 本の方程式群 $\mathbf{y}_{1i} = \mathbf{\Gamma}_1^{'} \boldsymbol{z}_{1i} + \mathbf{B}_2^{'} \boldsymbol{y}_{2i} + \mathbf{u}_i \ \ (i = 1, \cdots, n)$, (17) は $(K_1 + K_2) \times (G_1 + G_2)$ 分割行列

$$\begin{pmatrix} \mathbf{\Pi}_{11} & \mathbf{\Pi}_{12} \\ \mathbf{\Pi}_{21} & \mathbf{\Pi}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{G_1} \\ -\mathbf{B}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{\Gamma}_1 \\ \mathbf{0} \end{pmatrix} \tag{A.44}$$

で表現される。ここで構造方程式の誤差ベクトル $\mathbf{u}_i = (\mathbf{I}_{G_1}, -\mathbf{B}_2^{'}) \mathbf{v}_i = \mathbf{B}^{'} \mathbf{v}_i$、分散共分散が均一であれば $\mathbf{\Sigma} = \mathbf{B}^{'} \mathbf{\Omega} \mathbf{B},\ \mathbf{B}^{'} = (\mathbf{I}_{G_1}, -\mathbf{B}_2^{'})(G_1 \times (G_1 + G_2))$ である。一般に (33) は

$$\sqrt{n}\, \mathrm{vec} \begin{bmatrix} \hat{\mathbf{B}}_2 - \mathbf{B}_2 \\ \hat{\mathbf{\Gamma}}_1 - \mathbf{\Gamma}_1 \end{bmatrix} \xrightarrow{w} N[\mathbf{0}, (\mathbf{I}_{G_1} \otimes \mathbf{B}^{-1}) \mathbf{C} (\mathbf{I}_{G_1} \otimes \mathbf{B}^{-1})] \tag{A.45}$$

となる。ただし $\mathbf{A} \otimes \mathbf{B}$ は行列のクロネッカー積 $(a_{ij}\mathbf{B})$ を表し、

$$\mathbf{C}_{ij} = \mathrm{plim}_{n \to \infty} \begin{bmatrix} \mathbf{\Pi}_2^{'} \\ (\mathbf{I}_{K_1}, \mathbf{O}) \end{bmatrix} \frac{1}{n} \mathbf{Z}^{'} \hat{\mathbf{u}}_{i(n)} \hat{\mathbf{u}}_{j(n)}^{'} \mathbf{Z} \left[\mathbf{\Pi}_2, \begin{pmatrix} \mathbf{I}_{K_1} \\ \mathbf{O} \end{pmatrix} \right] ,$$

なおここで記号 $\hat{\mathbf{u}}_{i(n)}$ は第 i 番目の $n \times 1$ 残差ベクトル $(i = 1, \cdots, G_1)$ であり、$[G_1(G_2 + K_1)] \times [G_1(G_2 + K_1)]$ 行列 $\mathbf{C} = (\mathbf{C}_{ij})$ とした。
特に分散共分散が均一の時には $\mathbf{E}[\hat{\mathbf{u}}_{i(n)} \hat{\mathbf{u}}_{j(n)}^{'}] = \sigma_{ij} \mathbf{I}_n,\ \mathbf{\Sigma} = (\sigma_{ij})\ (i, j = 1, \cdots .G_1)$ となるので $\mathbf{C}_{ij} = \sigma_{ij} \mathbf{D}^{‘} \mathbf{M} \mathbf{D},\ \mathbf{C} = \mathbf{\Sigma} \otimes \mathbf{D}^{‘} \mathbf{M} \mathbf{D}$,

$$\mathbf{D} = \left[\mathbf{\Pi}_2, \begin{pmatrix} \mathbf{I}_{K_1} \\ \mathbf{O} \end{pmatrix} \right] ,\ (\mathbf{I}_{G_1} \otimes \mathbf{B}^{-1}) \mathbf{C} (\mathbf{I}_{G_1} \otimes \mathbf{B}^{-1}) = \mathbf{\Sigma} \otimes \mathbf{B}^{-1} .$$

となる。さらに $G_1 = 1$ とすると 4 節での表現に一致する。

[27]詳しくは Anderson, Kunitomo and Matsushita (2010) を参照されたい。