

SSE-DP-2025-6

二人の Wright 氏と操作変数法・構造方程式・
因果推論の潮流

池田真也(茨城大学農学部)・

国友直人(統計数理研究所)

2025 年 10 月

SSE-DP(ディスカッションペーパー・シリーズ)は以下のサイトから無料で入手可能です。

<https://stat-expert.ism.ac.jp/training/discussionpaper/>

このディスカッション・ペーパーは、関係者の討論に資するための未定稿の段階にある草稿である。著者の承諾なしに引用・複写することは差し控えられたい。

SSE-DP-2025-6

Two Professors Wrights and IV, Structural
Equations Model, and Statistical Causal Inferences

by

Shinya Ikeda and Naoto Kunitomo

The Institute of Statistical Mathematics

October 2025

(Summary)

Phillip Wright (1928) introduced a statistical estimation procedure, which can be regarded as IV (instrumental variables) method. Sewal Wright (1932) developed the path analysis. Since then, the structural equations model and simultaneous equations model have been developed in econometrics. In biometrics the IV method became popular since it is related to the noncompliance problem. We discuss these issues in historical as well as theoretical perspectives.

二人の Wright 氏と操作変数法・構造方程式・ 因果推論の潮流*

池田真也[†]・国友直人[‡]

2025 年 10 月

鍵言葉 (Key Words) : Phillip Wright と Sewall Wright, 操作変数法, 構造方程式モデル, 識別性, 同時方程式モデル, Noncompliance モデル, 統計的因果推論

要約

統計数理研究所が推進している統計エキスパート養成プログラムでは必ずしも統計学を専門としているわけではない各分野の若手研究者と統計家であるメンターにより共同研究が推奨されている。2025 年度に実施した共同研究の中から応用統計学の基礎と応用について二人の Wright 教授の古典的な研究に端を発してその後計量経済学、計量生物学をはじめ応用統計学の諸分野に影響が大きい「操作変数法・構造方程式・因果推論」を話題として検討したのでその成果を報告する。

*統計エキスパート養成事業における共同研究の報告書。

[†]茨城大学農学部

[‡]統計数理研究所

目次

はじめに

Part-I : 「操作変数法の起源 : アメリカ農業に関わる二人の Wright 氏の協働」 池田真也

Part-II : 「二人の Wright 氏と構造方程式・因果推論の潮流」
国友直人

おわりに

はじめに

統計数理研究所が推進している統計エキスパート養成プログラムでは必ずしも統計学を専門としているわけではない各分野の若手研究者と統計学分野を専門としているメンターにより統計エキスパート演習をおこなっている。2025年度には研修生の池田とメンターの国友が実施した共同研究では近年になり計量経済・生物統計など応用統計学の分野で関心が高まっている「操作変数法・構造方程式・統計的因果推論」を話題として検討したのでその成果を報告する。

統計エキスパート養成事業では全体研修という場で研修生は様々な話題についての模擬講義を行っている。元々、開発経済・農業経済を専攻している池田氏は2025年春の全体研修で計量経済学に関連が深い「操作変数法 (Instrumental Variables method)」について米国の一部の計量経済学者が最近になり注目している Phillip Wright (1928) の研究について報告を行った。

伝統的な計量経済学 (econometrics) においては操作変数 (Instrumental Variables, IV) という言葉は Olav Reiersol 教授が1945年に利用した研究 (学術誌「*Econometric Theory*, 16, 2000, 113-125, Cambridge UP」のインタビュー記事による) で初めて登場、その後 LSE (London School of Economics) に属した J. Durbin, D. Sargan 等による研究¹ により定着したと言う事情と理解されていた。したがって P. Wright (1928), S. Wright (1932) の研究を知らなかった年配の国友は改めて Phillip Wright, Sewal Wright が1920年代に行った研究をたどり、その後、Pearl (2009) などで議論されているパス解析などの内容を検討した。共同研究を進めている中で池田氏と国友が行った検討内容は操作変数法や構造方程式モデルや統計的因果推論の単なる歴史的経緯の事実関係だけでなく、現代の応用統計学の理解に有意義であるというように判断することになった。実は本論で説明するように近年での計量経済学、生物統計学などの応用につながる重要な論点が少なからずあるのである。そこで計量生物や計

¹Durbin (1954), "Errors in Variables," *Review of International Statistical Institute*, 22, 23-33. Sargan (1958), "The Estimation of Economic Relationship using Instrumental Variables," *Econometrica*, 26-3, 393-415.

量経済を含め応用統計学に関心のある統計エキスパートを読者として想定してここで Discussion Paper の一つとしてとりあげて報告することにした。

本稿の前半 Part I に収録した論文では池田氏が「操作変数法の起源：アメリカ農業に関わる二人の Wright 氏の協働」として米国の農業経済学における関税効果の検証というあまり知られていなかったかなり具体的問題から今日では”操作変数法”と呼べる統計的方法が考えられた経緯を説明する。後半の Part II に収録した論文では国友が「二人の Wright 氏と構造方程式・因果推論の潮流」として二人の Wright 教授が提起した統計的問題の理論的側面を加味して、その後の構造方程式モデルの利用と統計的因果推論についての潮流を考察した。

ここでとりあげた統計的問題に関心がある関係各位からのコメントを期待したい。

国友直人 (著者代表)

2025 年 10 月

操作変数法の起源：アメリカ農業に関わる二人の Wright 氏の協働

池田真也^{† ‡ *}

2025 年 10 月 4 日

概要

操作変数法を最初に発見したとされる Wright (1928) の業績は近年まで十分に認知されていなかった。Wright (1928) はアメリカの経済学者 Philip G. Wright の著書であるが、息子であり遺伝学者である Sewall Wright も操作変数法の発見に深く関わっており、Sewall による Wright (1934) では Wright (1928) の再検討が行われている。Sewall は現在の因果推論の潮流における DAG などのグラフィカルなアプローチの先駆者として知られており、Wright (1928) はいわば計量経済学と因果推論の交差点に位置する先駆的研究として評価できる。しかし、その重要性に比して Wright (1928) を解説した邦語文献は十分ではないように思われる。Wright (1928) ではモーメント法による推定とパス解析の応用による推定の 2 通りの方法でバター・亜麻仁の供給・需要の価格弾力性を求めており、本稿では Wright (1934) にも触れつつ解説する。

1 はじめに：Wright 親子の業績を見直す潮流

現在、構造方程式の識別において操作変数法の利用を最初に提示した業績として、Wright (1928) の Appendix B が参照される。そこではアメリカ農業において重要視され

[†] 茨城大学学術研究院応用生物学野

[‡] 統計数理研究所

^{*} shinyaik@gmail.com

ていたバターと亜麻仁の供給と需要の価格弾力性の推計方法が示されている。¹ところが、計量経済学分野においては当時から Wright (1928) が認知されていたわけではなく、2000 年代以降になってようやく定着したのである。

認知されるまでの経緯、つまり Wright (1928) を見直す潮流は示唆に富むものである。そもそも著者の Philip G. Wright は計量経済学の主流から外れていたこともあり、Goldberger (1972) による計量経済学および社会科学周辺分野における構造方程式の推計手法のサーベイの中での再発掘を待つことになった。さらに、その後も誰が操作変数法の発見者なのかという点で議論が残った。Wright (1928) の Appendix B ではモーメント法を用いた操作変数推定量とともに、パス解析を用いたグラフィカルな推定法も提示されており、後者の推定法に関しては、Philip G. Wright の息子であるとともに、遺伝学者およびパス解析の創始者として著名な Sewall Wright (米シカゴ大動物学科) によるものだとする考えがあったからである。Stock and Trebbi (2003) は過去の Philip G. Wright の著作を参照して文体計量分析 (Stylometric Analysis) により、Wright 1928 の Appendix B の執筆者が父親である Philip G. Wright だと結論付けている。しかし、2011 年の Kerry Clark によるシンポジウム報告では Philip と Sewall の間で交わされた 1925 年から 1926 年の書簡を検討しており、そこではまず息子の Sewall がパス解析による推定方法を提示したのち、父親の Philip がパス解析の知識がなくても可能な推計方法を提示したことが示されている。まさに親子の共同研究で操作変数法を開発したと言えるだろう。²その後、Sewall は Wright (1934) で Wright (1928) の議論を再検討している。そして、Sewall のパス解析は Judea Pearl の因果推論 (Pearl, 2009) へと繋がっているのである。

¹ 当時からバターは酪農業における重要な加工品であり、国内の農業政策において重要品目として認知されていた。また、亜麻仁 (flax seed) は亜麻の種子であり、抽出された油はオメガ 3 脂肪酸を豊富に含み、栄養価の高い油として広く消費されている。また、亜麻の茎はリネン繊維の原料であり、当時から重要な農産物加工品として認識されていた。

² Tafts 大学で行われた Philip G. Wright の生誕 150 周年記念シンポジウム (2011 年 10 月 3 日開催) における報告資料のアーカイブを参照されたい (<https://web.archive.org/web/20151210143523/http://ase.tufts.edu/economics/documents/wrightPhilipAndSewall.pdf>, 2025 年 9 月 27 日アクセス)。Wright (1928) 内では触れられていないが、書簡では小麦の需要の価格弾力性を操作変数法を用いて推計した結果についてのやりとりがあり、弱操作変数の問題まで検討していたことが明らかにされている。残念ながらその書簡資料を筆者は入手できなかったため再検証は叶わなかったが、当時から操作変数法を統計的に議論されていたことが伺える。

Wright (1928) と Wright (1934) は現在でも計量経済学、因果推論の分野においての先駆的研究であるが、具体的にどのように操作変数法が用いられたかを解説した日本語文献は管見の限り見当たらない。そこで本稿では、Wright (1928) を中心としつつ Wright (1934) も含めた議論を紹介することを目的とする。

2 Wright (1928) の概要

Wright (1928) は書籍全体を通じてアメリカの油脂関連製品に関する関税政策の影響を分析している。すなわち、アメリカにおける国内農林業生産者の保護を主目的とした 1921 年・1922 年の関税引き上げの有効性を動物油・植物油関連製品とその原料に着目して分析したのである。動物油・植物油の原料生産者が農産物価格の低さに苦しんでおり、その救済処置が背景にある。そのため、この関税の引き上げが原料生産者にとって有益なものなのか、また消費者は過大な負担を強いられないのか、という社会厚生 of 改善に目を向けた関税政策の検討が Wright (1928) の目的である (Wright, 1928, pp. 2–3)。結論としては、関税の引き上げの効果はどの動物油・植物油関連製品を例にとっても限定的であるとしており、否定的であった。

Wright (1928) の分析方法について触れると、データが不十分であった時代背景もあり、統計的な分析は限定的であった。たとえば、関税の多寡が価格、生産量、輸入量・輸出量に与える影響を検討するために、関税の変化の前後におけるこれら基本的データの変化を丁寧に考察していた。また、統計的手法は対象とする産業とその市場環境に対する深い知見と補完的に用いられるべきだとしており (Wright, 1928, p. 136)、複雑な動物油・植物油関連製品とその原料の生産過程、および製品間、原料間の代替性に関する定性的な解説と分析が採用されていた。³ バターを例に挙げれば (Wright, 1928, p. 17)、関税によりバターの国内価格が上昇した場合に酪農家は牛肉よりも牛乳の生産を重視し、その牛乳を飲料用やチーズなどの加工生産ではなく、バターの生産に当てるだろう。そうなるとバター

³ 本稿で取り上げる操作変数法に関する記述ではないため詳細は控えるが、当時のアメリカの油脂関連産業のフードシステムを理解するうえでの貴重な分析であり、その史料的な価値は十分に高い点は指摘しておきたい。

の国内生産は増加して価格は低下する傾向が生じるため、関税引き上げの目的は無効化されてしまう。Wright (1928) の第 6 章では動物油・植物油を併せた脂肪油のなかでも、酪農産業の主力加工品であり、また関税政策上重要性が高いバターのみを取り上げ、その関税の変化の影響を検討している。

この第 6 章では、Wright (1928) の Appendix B で操作変数法を用いて推定したバター需要の価格弾力性を用いて議論が展開されている (Wright, 1928, pp. 147–150)。主なバターの輸入先であるコペンハーゲンのバターの価格とニューヨークのバターの価格の差が関税によりどの程度変動したかを需要の価格弾力性を用いて推計しているのである。具体的には実際の 12.5 セントの関税が仮想的に 2.5 セントの以前からの課税水準だった場合に、国内消費に回るバターが約 1.4 パーセント増加したと計算される（輸入の変動分が国内消費に回るとし、その輸入の変動は実際の 12.5 セントの関税時の輸入量から 2.5 セントの時の輸入量の差分として計算している）。バターの需要の価格弾力性が約-0.6 なので、価格の減少幅は 2.3 パーセントである。期間中のニューヨークのバターの平均価格が 47.4 セントであるため、価格は約 1.1 セント下落したはずである。つまり、関税を 10 セント引き上げても 1.1 セントしか価格は増加しなかったことになり、当時一般に考えられていたよりも関税の影響は小さい点が述べられている。

なお、Wright (1928) の Appendix B ではバターの供給と需要の価格弾力性が推定されているが、同時に亜麻仁についても弾力性が推計されている。亜麻仁は植物油の主要な原料である点や、バターとは異なり非必需品であり、需要弾力性についての良い比較対象になることが理由とされている (Wright, 1928, p. 1, pp.318–319)。

3 操作変数法による価格弾力性の推計：Wright (1928)Appendix B における 2 通りの方法

Wright (1928) の Appendix B では線形構造方程式により需要関数と供給関数をモデル化している。⁴通常は価格系列は時系列データになるため下添え字を年度 t とし、価格を

⁴ 本節での Wright (1928) のモデルを紹介するうえで、Abbring et al. (2025) の解説を参考にした。

P_t とする。需要量 $D_t(p_t)$ について外生要因である U_d を加えたものが式 (1) である。同様に供給量 $S_t(p_t)$ とその外生要因 U_s を加えたものが式 (2) である。ここで、 $D_t(p_t)$ と $S_t(p_t)$ が一致する均衡状態のときの価格 P_t と量 Q_t が実際には観察されるため、 $D_t(p_t)$ と $S_t(p_t)$ は観察できない変数である点に注意されたい。

$$D_t(p_t) = \alpha_1 p_t + U_t^d \quad (1)$$

$$S_t(p_t) = \beta_1 p_t + U_t^s \quad (2)$$

$$D_t(p_t) = S_t(p_t) = Q_t \quad (3)$$

それぞれの観測可能な変数について、たとえば価格 P_t であれば名目値 R_t を使うのではなくパーセンテージ偏差 (percentage deviation) $\frac{R_t - \mathbb{E}[R_t]}{\mathbb{E}[R_t]}$ に変換したうえで取り扱っており、 $\mathbb{E}[R_t]$ は R_t の観測期間の平均値として計算している。そのため、Wright (1928) では U_t^d と U_t^s は平均 0 で互いに相関しない確率的なショックと同様に扱われている。また、パラメータ α_1 に着目すると、価格 p_t が 1 パーセント変化したことに対する需要量 D_t の変化率を表すため、需要の価格弾力性を表している。同様に、 β_1 は供給の価格弾力性と一致する。

α_1 と β_1 を推定するにあたって、均衡状態の価格 P_t と量 Q_t の 1 組しか観察することができないが、そこから需要曲線と供給曲線をそれぞれ区別するためには、需要曲線と供給曲線のどちらかだけが変動していることが分かっていることが必要だと Wright (1928) は考察している。一般的にはそのような状況の方が珍しいため、『... statistical methods for imputing fixity to one of the curves while the other changes its position must be based on the introduction of additional factors. Such additional factors may be factors which (A) affect demand conditions without affecting cost conditions or which (B) affect cost conditions without affecting demand conditions. (一方の曲線を固定し、もう片方の曲線が位置を変えるようにするための統計的方法では、追加的ファクターの導入に頼らざるを得ない。そのような追加的ファクターとしては、(A) 需要条件に影響するが費用条件に影響しない要因、または (B) 費用条件に影響するが需要条件には影響しない要因がありうる。)] (Wright, 1928, pp. 311–312) と述べている。つまり、需要（供給）関数

に対する外生的な変動要因（供給（需要）シフターとも呼ばれる）を利用するアイデア、すなわち操作変数の利用を提示していたと解釈できる。

操作変数の例としては、供給関数にしか変動を与えない要因として1エーカーあたりの収穫量いわゆる単収と、前年度の価格を例として挙げており、バターと亜麻仁を用いた推計の際には単収を主な操作変数の例として挙げている (Wright, 1928, p. 312)。少なくとも亜麻仁の需要の価格弾力性を推計する際には6つの異なる供給シフターを操作変数に使用し、その平均値から-0.8と推計している (Wright, 1928, p. 319)。4つの推計値は1を下回ったが、2つは1を上回ったとあり、亜麻仁の需要の価格弾力性についてはかなり幅が広いことが伺える。実際の推計に用いられた6つの供給シフターは特定されていないが、生産量を作付け面積と単収に分解し、単収は天候に依存する点、作付け面積は亜麻仁の代替的作物である春小麦との前年度における単位面積当たりの相対的収益差に依存する点が触れられており (Wright, 1928, p. 195)、これらが操作変数の候補であったことが伺える。

ここで、操作変数を式 (1)(2) に含めるようにすると、 U_t^d について、需要にのみ影響して供給には影響しない要因を Z_t^d とおく。また同様に供給のみに影響して需要には影響しない要因を Z_t^s とおき、一般には共通要因 W_t と誤差項により下記のように定式化される。

$$U_t^d = \alpha_2 Z_t^d + W_t + \epsilon_t^d \quad (4)$$

$$U_t^s = \beta_2 Z_t^s + W_t + \epsilon_t^s \quad (5)$$

以降では Wright (1928) と同様にするため、共通要因 W_t は無いものとし、また Z_t^d と Z_t^s の間は無相関を仮定する。なお、添え字の t は可読性のために以後省略する。

価格弾力性の推計のために使用されたデータに関して、Wright (1928) では明示されていないが、1920年から1926年までの価格と消費量のデータを使ったと推測される。しかし、Wright (1928) 内には年別のデータしか示されておらず、実際にどのようなデータを使用したのか、たとえば月別のデータの利用の有無は確かめることができなかった。⁵

⁵ Wright (1928) から関連する統計データを引用して、前年度価格を操作変数としてバターの需要の価格弾力性の試算を行った。年別のデータを用いた場合は、前年度比のデータを用いる都合上サンプルサイズは5と非常に少なくなるわけだが、実際に価格弾力性が2.2と推計され、Wright (1928) で示された-0.6とは大きく異なった。筆者はさらに当時の月別データの使用可能性を検討した。貿易関連

3.1 操作変数推定量

Wright (1928) では、図 1 のように原点を通る供給曲線 S と需要曲線 D がそれぞれ U^s , U^d のショックを与えられ S' と D' に移動した状況をグラフィカルに考察した。⁶ 前述のように価格と量に関するデータはパーセンテージ偏差が取られているため U^s と U^d は平均 0 であり、そのため供給曲線 S と需要曲線 D は平均的には原点 o を通る。そして、 S' と D' における水平軸との交点がそれぞれ $(U^s, 0)$ 、 $(U^d, 0)$ であり、また S' と D' の交点を (P, Q) とする。需要の価格弾力性 α_1 は需要関数の垂直軸から見た傾きに等しいことから、

$$\alpha_1 = \frac{Q - U^d}{P} \quad (6)$$

となる。その後、式 (6) の両辺に Z^s を掛ければ、 $\alpha_1 Z^s P = Z^s Q - Z^s U^d$ であり、標本

項目や価格のデータは問題なく月別のデータが存在する。問題となるのは消費量のデータである。実は Wright (1928) は直接的に消費量のデータを持っていたわけではなく、生産量・輸出量・輸入量のデータから消費量を推計している (Wright, 1928, p. 267)。その際の生産量のデータのソースとして、(Wright, 1928, p. 261) には 1919-1926 年までについては米商務省統計局 (Bureau of the Census, U.S. Department of Commerce) とあるが、年別のデータであっても現在手に入る生産量のデータとは異なる。たとえば、Wright (1928) では 1923 年のバターの国内生産量は 1862 (百万ポンド) であるが、別のリソースでは 1986 (百万ポンド) である (United States. Bureau of the Census. *Historical Statistics of the United States, 1789-1945: A Supplement to the Statistical Abstract of the United States*. Washington, D.C.: U.S. Government Printing Office, 1949; https://www.census.gov/library/publications/1949/compendia/hist_stats_1789-1945.html)。これはバター生産の 3 割程度が工場製ではなく家庭製であり、そのデータの把握を粗い推計に頼らざるをえないことが要因であろう。また、現在では工場製のバター生産に関してのみ月別の長期時系列データの入手は NBER のバターの生産量データベース (National Bureau of Economic Research, Butter Production in Factories for United States [M01069USM149NNBR], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/M01069USM149NNBR>, September 26, 2025.) から利用可能だが、家庭製まで含めた月別の生産データの入手は困難である。さらに、Wright (1928) ではバターの貯蔵分が考慮されていないため、現在の NBER のバターの消費データ (National Bureau of Economic Research, Butter Consumption for United States [M01071USM149NNBR], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/M01071USM149NNBR>, September 26, 2025.) と比較するとかなり過大に消費量を推計していると考えられる。

⁶ 本稿では触れないが、仮にある時点 t から $t+1$ にかけて供給か需要のどちらかのショックしか生じていないと仮定すれば、例えば供給曲線のみ変動したとすれば、価格と量の変化の向きは互いに異なるようになる。一方で需要関数のみ変動したとすれば、価格と量は同じ方向に変化することになる。この点を利用して、データを分類して弾力性の推定を行った例も提示している (Wright, 1928, 316)。

全体を足し合わせれば、

$$\alpha_1 = \frac{\sum(Z^s Q) - \sum(Z^s U^d)}{\sum(Z^s P)} \quad (7)$$

となる。ここで $\sum(Z^s U^d)$ に注目すると、需要ショックにあたる U^d と供給にのみ影響を与える Z^s の相関は 0 になるため、 $\sum(Z^s U^d) = 0$ になる。従って、

$$\alpha_1 = \frac{\sum(Z^s Q)}{\sum(Z^s P)} \quad (8)$$

と推定している。これは現在では操作変数推定量 (instrumental variable estimator) として知られている方法である。Wright (1928) では U^d を確率変数と明記してはいないが、 Z^s と U^d が無相関である点をモーメント条件 (直交条件) として書き換えれば、式 (6) を用いて $\mathbb{E}[(Q - \alpha_1 P)Z^s] = 0$ と表すことができる。そして、操作変数推定量として知られる

$$\alpha_1 = \frac{\mathbb{E}[Z^s Q]}{\mathbb{E}[Z^s P]} \quad (9)$$

を得ることができるのである。

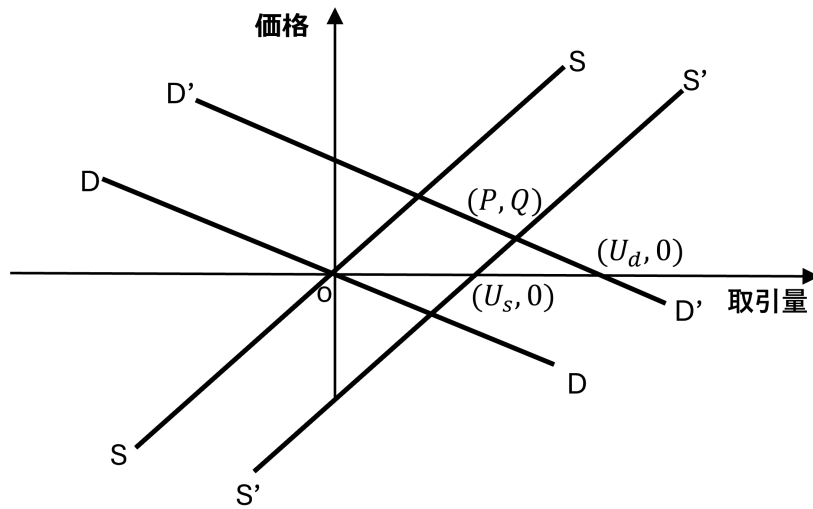


図 1: Wright (1928) が想定した需要曲線と供給曲線
出所: Wright (1928) Appendix B の図 9 を参照して筆者作成

3.2 パス解析の利用

ここでまず、式 (1)(2)(4)(5) をパス図で表現すると図 2 のようになる。次に式 (1)(2)(3) を用いて、 P と Q について解き、それをパス係数 p_1, p_2, q_1, q_2 と各変数の標準偏差 $\sigma_P, \sigma_Q, \sigma_{U^d}, \sigma_{U^s}$ で表現すると、式 (10)(11) となる。

$$P = \frac{1}{\beta_1 - \alpha_1}(U^d - U^s) = p_1 \frac{\sigma_P}{\sigma_{U^d}} U^d + p_2 \frac{\sigma_P}{\sigma_{U^s}} U^s \quad (10)$$

$$Q = \frac{\alpha_1}{\beta_1 - \alpha_1}(U^d - U^s) + U^d = q_1 \frac{\sigma_Q}{\sigma_{U^d}} U^d + q_2 \frac{\sigma_Q}{\sigma_{U^s}} U^s \quad (11)$$

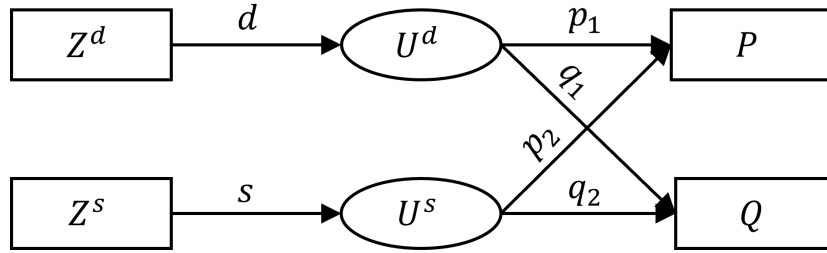


図 2: Wright (1928) の構造方程式のパス図表現
出所: Abbring et al. (2025) の図 4 を参照して筆者作成

ここで需要の価格弾力性 α_1 は、供給ショックのみが生じた場合に推計可能となるため、逆に言えば $U^d = 0$ の場合に推計されることから、式 (12) のように推計されることが分かる。供給の価格弾力性も同様にすれば、式 (13) のように推計される。

$$\alpha_1 = \frac{Q}{P} = \frac{q_2 \sigma_Q}{p_2 \sigma_P} \quad (12)$$

$$\beta_1 = \frac{q_1 \sigma_Q}{p_1 \sigma_P} \quad (13)$$

U^d と U^s は観測できない変数だったため、価格弾力性 α_1 と β_1 を推計するためには、未知のパス係数である p_1, p_2, q_1, q_2 を推計する必要がある。その際、パス解析では $r_{Z^s P} = p_2 s$ 、 $r_{Z^s Q} = q_2 s$ となることを利用すれば p_2 と q_2 を消去することができ、 $\alpha_1 = \frac{r_{Z^s Q} \sigma_Q}{r_{Z^s P} \sigma_P}$ となる。 Q と Z^s の相関係数はデータから $r_{Q Z^s} = \frac{\sum Q Z^s}{n \sigma_Q \sigma_{Z^s}}$ と計算されるので、

$$\alpha_1 = \frac{\sum(Z^s Q)}{\sum(Z^s P)}$$

と式 (8) と同一の結果となる。 β_1 についても同様に計算することが可能である。結果として前述の操作変数推定量と一致する。なお、内生変数 P と Q を外生変数に回帰させることで得られる誘導形の OLS 推定量を用いると、その比の形で式 (6) を得ることができる。これは間接最小二乗法 (indirect least squares method) と呼ばれる方法である (Abbring et al., 2025; Hansen, 2022, pp. 341–352)。⁷

Wright (1934) ではより詳細な検討が行われている。まず、 U^d と U^s の相関 $r_{U^d U^s} \neq 0$ を許容しつつ、操作変数 Z^d, Z^s を含まないモデルから出発している。未知のパス係数を推計するために、 P および Q の自己相関、 P と Q の相関、そして操作変数と $P \cdot Q$ との相関をパス解析を用いて 7 本の条件式から、7 つの未知パラメータ ($p_1, p_2, q_1, q_2, d, s, r_{U^d U^s}$) を識別している。ただし、Wright (1934) の実証分析では、操作変数が Z^s のみで、操作変数 Z^d をモデルに含めない代わりに $r_{U^d U^s} = 0$ を仮定している。つまり、5 本の条件式から 5 つの未知パラメータ (p_1, p_2, q_1, q_2, s) を求めているのである。⁸

1871 年から 1915 年までの食用豚の価格—取引量データを用いた例では、前年度のとうもろこし生産量等を生産にのみ影響する操作変数としており、生産の価格弾力性は 0.133、需要の価格弾力性は -0.944 という結果であった。また、1896 年から 1914 年までのジャガイモの価格—取引量データを用いた分析も行っており、その際は前年度のジャガイモ価格を生産にのみ影響する操作変数としていた。その生産の価格弾力性は 0.034、需要の価格弾力性は -0.815 であった。Wright (1934) は生産の価格弾力性が極端に低く、非弾力的で

⁷ 式 (10) と式 (11) において、 $U^d = 0$ の時に外生変数は Z^s のみであり、内生変数を外生変数に回帰させれば、 $P = -\frac{\beta_2}{\beta_1 - \alpha_1} Z^s + u^p$ 、 $Q = -\frac{\alpha_1 \beta_2}{\beta_1 - \alpha_1} Z^s + u^q$ を得る (ここで、 $u^p = -\frac{1}{\beta_1 - \alpha_1} \epsilon^s$ 、 $u^q = -\frac{\alpha_1}{\beta_1 - \alpha_1} \epsilon^s$ である)。得られた OLS 推定量の比が α_1 になることは容易に確かめられる。 β_1 についても同様であり、Wright (1928) のように内生変数と操作変数の数が同じ場合 (丁度識別; just-identified) に用いることができる手法である。

⁸ この議論を Wright (1928) に適用すると過剰識別をしていたことになり、パラメータの識別ができていなかった可能性が気になる。実際のところ、 $p_1 = \pm \sqrt{\frac{1-l^2}{k^2-l^2}}$ と $p_2 = \pm \sqrt{\frac{k^2-1}{k^2-l^2}}$ となり (ただし、 $k = \frac{r_{Z^d Q}}{r_{Z^d P}}$ 、 $l = \frac{r_{Z^s Q}}{r_{Z^s P}}$)、一意に定まらない。しかし、 $\frac{q_2}{p_2} = \frac{r_{Z^s Q}}{r_{Z^s P}}$ と計算されるため、結果には影響しない。なお、Wright (1928) において $r_{U^d U^s} \neq 0$ を許容しても計算は多少複雑になるが $\frac{q_2}{p_2}$ の計算には影響しないことが確かめられる。

あった原因が単年度の需給調整しかモデルで考慮していないからとしている。すなわち、前年度の価格を指標として翌年度の農家の生産行動が強く決定されており（推計されたパス係数 $s = 0.651$ ）、長期の平均価格と平均取引量という観点で見れば価格変化に対して弾力的になっている可能性を指摘した。そのうえで、再帰的なモデルを考察している（図 2 のモデルが各年（期）に設定され、年度 t と年度 $t+1$ のパス図が P_t から Z_{t+1}^s へのパスで結ばれ、全ての期が連結しているモデル（パス図）である。関心のある読者は Wright (1934) の図 32 を参照してほしい）。

4 おわりに

本稿では Wright (1928) において示された操作変数法のアイデアを紹介した。一般的な計量経済学の教育現場で紹介されるのはモーメント条件を利用した操作変数推定量の方である。そして、Wright (1928) では最初に操作変数推定量のアイデアが提示され、より完全な方法（a more complete analysis (Wright, 1928, p. 315)）として Sewall のパス解析に言及している。しかし、当時の資料によれば、操作変数法の起源においては Sewall Wright によるパス解析の観点からの識別問題への洞察が重要な役割を果たしており、その代替的方法として父親の Philip G. Wright が操作変数推定量のアイデアを提示したという点は興味深い。

また、操作変数法を含めた計量経済学の手法は、何らかの現実の問題から生まれることが多いが、操作変数法が適用された問題が農産物（加工品）市場に対する輸入関税の影響だったことは Wright (1928) の業績を紹介する既存文献であまり触れられていなかった点である。2025 年時点でトランプ関税が議論される現状を鑑みるに、1 世紀経っても色褪せない問題を取り上げていたわけである。本稿では Wright (1928) でどのような分析が行われたかを簡易的に紹介したにすぎないが、少なくとも当時のアメリカの油脂関連産業の実態を理解するという意味においても Wright (1928) は価値があると思う。

参考文献

- Abbring, Jaap H., Victor Chernozhukov, and Iván Fernández-Val**, “Philip G. Wright, directed acyclic graphs, and instrumental variables,” *Econometrics Journal*, 2025, 28 (1), 1–20.
- Goldberger, Arthur S**, “Structural Equation Methods in the Social Sciences,” *Econometrica*, 1972, 40 (6), 979–1001.
- Hansen, Bruce E.**, *Econometrics*, Princeton University Press, 2022.
- Pearl, Judea**, *Causality*, Cambridge University Press, 2009.
- Stock, James H and Francesco Trebbi**, “Retrospectives: Who Invented Instrumental Variable Regression?,” *Journal of Economic Perspectives*, 2003, 17 (3), 177–194.
- Wright, Philip Green**, *The Tariff on Animal and Vegetable Oils* Investigations in international commercial policies, Macmillan, 1928.
- Wright, Sewall**, “The Method of Path Coefficients,” *The Annals of Mathematical Statistics*, 1934, 5 (3), 161–215.

二人の Wright 氏と構造方程式・因果推論の潮流¹

国友直人²

2025 年 9 月 29 日

鍵言葉 (Key Words)

二人の Wright 教授, 構造方程式と同時方程式, DAG モデル, 非遵守モデル, 操作変数法, 構造方程式モデルの制限情報推定法, 統計的因果推論, Noncompliance モデル

概要 (Summary)

本稿では構造方程式モデル・因果推論・操作変数法の源流として経済学者の研究 Wright (1915), 遺伝学者の研究 Wright (1921) まで遡る。計量経済学における伝統的な同時方程式モデルを巡る議論を利用して応用統計学における構造方程式モデルとグラフ理論における DAG(directed acyclic graph) モデルの制約を特に Wold(1954) の Causal Chain モデルを通じて説明する。さらに応用例として計量生物における Noncompliance (非遵守) モデルにおける操作変数法の利用をとりあげ、計量経済学、生物統計学、応用統計学における統計的因果推論の展開を展望する。

1 はじめに

「観測変数間の因果関係を線形回帰モデルで記述したものをパス解析という。1930 年代に生物学者の Sewall Wright によって創始されたパス解析はその後認知心理学者... によって研究され、計量経済学の同時方程式モデル、そしてここで議論している構造方程式モデルへと発展した。...」これはしばらく前に日本で出版された書籍（統計科学のフロンティア vol.5, 狩野裕「構造方程式モデリング, 因果推論, そして非線形性」, Page 73, 岩波, 2002）に書かれている記述であり、日本の応用統計学ではかなり流布されている議論と思われる。

¹2025-9-29. 統計エキスパート slack 上で行われた池田真也氏との共同研究の一環でとりあえずまとめた未完成的な原稿であり、コメントを歓迎する。

²統計数理研究所, 〒 190-8562 東京都立川市緑町 10-3

しかしながら池田 (2025) が指摘しているように、経済学者 Phillip Wright (1928) は農産物の価格への関税の経済効果の計測問題から市場価格と数量から需要の弾力性を計測しようとするときに生じる同時方程式における識別問題を指摘、需要の弾力性の推定にある種の操作変数法 (instrumental variables method) と (今では) 解釈できる統計的推定法を提案したのである³。したがって、引用した書籍について少なくとも最初の記述は正確ではないことが分るが、それでは計量経済学に関する記述はどうであろうか？少なくともしばらく前の米国で計量経済学を学んだ筆者にはその記述にはかなりの違和感がある。

例えば一例に過ぎないが、しばらく前の計量経済学の分野では構造方程式モデル・同時方程式モデルの利用は Haavelmo (1943, 1944) により提唱、当時の Cowles Commission (コールズ委員会) に属した統計家・計量経済学者により統計的推定法が開発されたことが常識とされていた。また操作変数 (instrumental variables) という名前は O. Reiersol が 1945 年に書いた論文に始まるとされていた⁴。

他方、冒頭の引用で紹介したように、日本の応用統計学ではこうした計量経済における議論とは関連することなく構造方程式モデルや DAG モデルの利用が盛んに行われている。さらに計量生物学では近年になり米国における議論を媒介として、操作変数法が利用されるようになっている⁵。

本稿では構造方程式モデルを巡る議論の展開の歴史的経緯を経済学者 P. Wright, 遺伝学者 S. Wright の研究まで遡るとともに、理論的な観点から、構造方程式モデル、同時方程式モデル、識別問題、DAG モデル、操作変数法、統計的因果推論という近年の統計学の利用において話題となっている基礎的事項を扱う。特に幾つかの簡単な例を用いて計量経済学、計量生物学、応用統計学などの異なる幾つかの分野でしばしば別々に議論されている統計的方法が統一的に理解できることを主張する。本稿では分かりやすい例を用いるが、問題の説明はより一般的な議論を展望する

³統計学の歴史では例えば竹内啓「歴史と統計学」Page120 には Mayer(1723-1762) は「変数誤差の統計的処理で操作変数と解釈できる方法を利用した」ことを指摘している。

⁴学術誌「*Econometric Theory*, 16, 2000, 113-125」に掲載されたインタビュー記事による。Haavelmo による同時方程式モデルの重要性の指摘を受け、コールズ委員会 (Cowles Commission) の A.Wald, T.Koopmans, T.W.Anderson などにより推測理論などが開発された。当時の事情については例えば Koopmans (1953) が詳しい。

⁵例えば岩崎 (2015) を参照されたい

のでいずれの例も、さらに現実的で複雑な統計モデルに拡張することは容易である。

本稿の以下は次のような内容で構成される。第二節では線形同時方程式 (linear simultaneous equations) モデルと線形構造方程式 (linear structural equations) モデルを Anderson(1979) にしたがって説明する。第三節では構造方程式モデルと DAG (directed acyclic graph) モデルの関係を議論する。第四節では応用例として計量生物における Noncompliance (非遵守) モデルをとりあげて操作変数法などの統計的推定問題を説明する。第五節では暫定的結論を述べる。さらに数理的付論では補論として証明と関連事項を説明する。

2 構造方程式モデルとは？

構造方程式モデルの利用を巡る幾つかの議論は「構造方程式とは何か？」という基本的な問題についての理解が異なる研究分野により異なることから生じるようである。ここではもっとも一般的と見られる計量経済分野における同時方程式モデルという観点から Anderson (1979) にしたがって同時方程式モデルと構造方程式モデルを扱ってみる。

クロスセクションのある個体、あるいは時刻を添字 t として多次元の従属変数 (内生変数) ベクトルと先決変数 (外生変数とラグ付き内生変数 $\mathbf{y}_{t-s}$ ($s \geq 1$)⁶) により

$$\mathbf{y}_t = \begin{pmatrix} y_{1t} \\ \vdots \\ y_{Gt} \end{pmatrix}, \mathbf{z}_t = \begin{pmatrix} z_{1t} \\ \vdots \\ z_{Kt} \end{pmatrix} \quad (1)$$

としよう。このとき変数間の関係を表現する線形構造方程式 (linear structural equations) を

$$\mathbf{B}\mathbf{y}_t + \mathbf{\Gamma}\mathbf{z}_t = \mathbf{u}_t \quad (2)$$

⁶ここでは時点 t においてあらかじめ値が決まっている変数、時刻 $t-1$ での σ -集合族が $\mathcal{F}_{t-1}$ のとき $\mathcal{F}_{t-1}$ 可測とする。外生性 (exogeneity) については後で議論するように幾つかの論点がある。

と表そう。ただし

$$\mathbf{u}_t = \begin{pmatrix} u_{1t} \\ \vdots \\ u_{Gt} \end{pmatrix}. \quad (3)$$

は構造方程式に現れる観測不能な誤差項であり、 $\mathbf{B} = (\beta_{gh})$, $\mathbf{g}, \mathbf{h} = 1, \dots, G$, $\mathbf{\Gamma} = (\gamma_{gk})$, $\mathbf{g} = 1, \dots, G$, $\mathbf{k} = 1, \dots, K$ は未知母数を表す $G \times G, G \times K$ 行列とする。この方程式群は G 個の内生変数 y_{gt} ($g = 1, \dots, G$), K 個の先決変数 z_{kt} ($k = 1, \dots, K$) を含み、(経済構造を表現する) 構造方程式 と呼ぶが、先決変数を所与とした内生変数の挙動を表現している。

ここで次のような条件を仮定する。

仮定 1: 行列 $\mathbf{B}$ は正則である。

仮定 2:

$$\mathbf{E}(\mathbf{u}_t) = \mathbf{0}, \quad (4)$$

$$\mathbf{E}(\mathbf{u}_t \mathbf{u}_t') = \Sigma (= \sigma_{gh}). \quad (5)$$

仮定 3: 誤差ベクトル $\mathbf{u}_t$ は (t について) 互いに独立とする。

仮定 4: 先決変数ベクトル $\mathbf{z}_t$ の要素は外生的 (exogenous)⁷、あるいは内生変数の過去値 $\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots$ とする。

仮定 5: 内生変数の初期ベクトル $\mathbf{y}_0, \mathbf{y}_{-1}, \dots$ は所与、先決変数ベクトル $\mathbf{z}_1, \mathbf{z}_2, \dots$ の要素は観測可能な変数とする。

仮定 1 から構造方程式群の左から $\mathbf{B}^{-1}$ を乗ずれば

$$\mathbf{y}_t = -\mathbf{B}^{-1}\mathbf{\Gamma}\mathbf{z}_t + \mathbf{B}^{-1}\mathbf{u}_t, \quad (6)$$

が得られる。行列 $\mathbf{\Pi} = -\mathbf{B}^{-1}\mathbf{\Gamma}$, 確率ベクトル $\mathbf{v}_t = \mathbf{B}^{-1}\mathbf{u}_t$ と置くと、誘導形

$$\mathbf{y}_t = \mathbf{\Pi}\mathbf{z}_t + \mathbf{v}_t \quad (7)$$

が得られる。誘導型の母係数 $\mathbf{\Pi} = (\pi_{gk})$, ($g = 1, \dots, G, k = 1, \dots, K$) のとき誤差項の期待値 $\mathbf{E}(\mathbf{v}_t) = \mathbf{E}(\mathbf{B}^{-1}\mathbf{u}_t) = \mathbf{0}$, 分散・共分散行列

$$\mathbf{\Omega} = \mathbf{E}(\mathbf{v}_t \mathbf{v}_t') = \mathbf{B}^{-1}\mathbf{\Sigma}\mathbf{B}^{-1'} \quad (8)$$

および $\mathbf{E}(\mathbf{v}_t \mathbf{v}_s') = \mathbf{E}(\mathbf{B}^{-1}\mathbf{u}_t \mathbf{u}_s' \mathbf{B}^{-1'}) = \mathbf{0}$ ($t \neq s$) となる。

構造方程式の誤差ベクトルの G 次元の確率密度 $f(\mathbf{u}_t)$ 、誤差確率ベクトル

⁷外生性 (exogeneity) については再び議論するが、システムの外からあらかじめ決まっている (確率 or 非確率) 変数とする。

が $t = 1, \dots, T$ について互いに独立とすると、観測される確率変数ベクトル $\mathbf{y}_1, \dots, \mathbf{y}_T$ の ($\mathbf{z}_t$ ($t = 1, \dots, T$)) を条件とした) 同時分布は

$$g(\mathbf{y}_1, \dots, \mathbf{y}_T | \mathbf{z}_1, \dots, \mathbf{z}_T) = \text{mod}|\mathbf{B}|^T \prod_{t=1}^T f(\mathbf{B}\mathbf{y}_t + \mathbf{\Gamma}\mathbf{z}_t). \quad (9)$$

で与えられる。ただし $\text{mod}|\mathbf{B}|$ は $\mathbf{u}$ から $\mathbf{v}$ への変換のヤコビアン (行列式の絶対値) であり、各 t に対して $\partial u_{gt} / \partial y_{ht} = b_{gh}$ (ただし $\mathbf{B} = (b_{gh})$) より $|\mathbf{B}|$ の絶対値、観測全体のヤコビアンは $|\mathbf{B}|^T$ の絶対値になる。なお統計的分析でよく利用される多次元正規分布の場合にはかなり簡単な表現になる。

例 1: 競争的な取引で価格と取引量が決まるある商品の市場を考える。時刻 t における商品の価格を p_t 、取引量を q_t とする。需要関数 (例えば消費者の需要 $q_t^{(d)}(p_t)$) と供給関数 (例えば生産者の供給 $q_t^{(s)}(p_t)$) が線形近似できるとしてそれぞれ

$$\text{D: } q_t^{(d)} = \alpha + \beta p_t + \epsilon_t, \quad (10)$$

$$\text{S: } q_t^{(s)} = \gamma + \delta p_t + \eta_t \quad (11)$$

で表現してみよう⁸。ここで $\alpha, \beta, \gamma, \delta$ は (ゼロでない) 未知母数、 ϵ, η は誤差項とする⁹。この時にこの商品市場において需要と供給が均衡状態 ($q_t = q^{(d)} = q^{(s)}$) で価格が決定され、価格と数量が観測されるとすると

$$\mathbf{y}_t = \begin{pmatrix} q_t \\ p_t \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & -\beta \\ 1 & -\delta \end{pmatrix}, \quad \mathbf{z}_t = \mathbf{1}, \quad \mathbf{\Gamma} = \begin{pmatrix} -\alpha \\ -\gamma \end{pmatrix} \quad (12)$$

と表現される。この構造方程式モデルは $G = 2$, $\mathbf{B}$ が三角行列ではなく、 $\mathbf{\Sigma}$ は対角行列とは限らない同時方程式 (simultaneous equations) モデルの例である。誘導型は ($\delta - \beta \neq 0$ を仮定すると) 価格と数量 p_t, q_t について

$$p_t = \frac{\alpha - \gamma}{\delta - \beta} + \frac{\epsilon_t - \eta_t}{\delta - \beta} = \phi + u_t, \quad (13)$$

⁸19 世紀の後半に起きた限界革命 (marginal revolution と呼ばれている) から欧米の主流派経済学では需要と供給が一致する均衡で価格と数量が決まるとする考え方が基本的であり、株価など金融市場を含め近年では日本の経済系の大学教育もこの考え方で行われている。不均衡理論と統計モデルも存在するが一般的ではないのでここでは議論を省略する。

⁹なお誤差項同士の相関は一般にはゼロとは仮定しない。ゼロとは考えられない例としては、供給が鉄鋼メーカー、需要が自動車メーカー、ノイズはマクロ経済ショックなどの例を考えることができる。この論点は後述の DAG による構造方程式モデルの表現可能性と関係する。

$$q_t = \frac{\delta\alpha - \beta\gamma}{\delta - \beta} + \frac{\delta\epsilon_t - \beta\eta_t}{\delta - \beta} = \theta + v_t, \quad (14)$$

となる。ただし $\phi = (\alpha - \gamma)/(\delta - \beta)$, $u_t = (\epsilon_t - \eta_t)/(\delta - \beta)$, $\theta = (\delta\alpha - \beta\gamma)/(\delta - \beta)$, $v_t = (\delta\epsilon_t - \beta\eta_t)/(\delta - \beta)$ である。ここで誤差項が二次元正規分布にしたがい分散・共分散 $\sigma_\epsilon^2, \sigma_\eta^2, \sigma_{\epsilon\eta}$ とすると条件付期待値は

$$\mathbf{E}(q_t|p_t) = \theta + \left[\frac{\beta\sigma_\eta^2 + \delta\sigma_\epsilon^2 + (-\beta - \delta)\sigma_{\epsilon\eta}}{\sigma_\eta^2 + \sigma_\epsilon^2 - 2\sigma_{\epsilon\eta}} \right] (p_t - \phi)$$

となる¹⁰。したがって取引量データ q を価格データ p に回帰すると最小二乗推定では必ず推定バイアスが発生する。例えば $\sigma_{\epsilon\eta} = 0$ のとき σ_ϵ が小さければ β の推定バイアスは小さくなるが、 σ_η が小さければ δ の推定バイアスが小さくなる。またここでの構造方程式に登場する母数は $\theta = (\alpha, \beta, \gamma, \delta, \sigma_\epsilon^2, \sigma_\eta^2, \sigma_{\epsilon\eta})'$ の7個、誘導型の母数は5個である。

例 2: 需要-供給の市場モデルにおいて供給関数が過去の価格に依存する動学モデルの簡単な例を

$$\text{D: } q_t^{(d)} = \alpha + \beta p_t + \epsilon_t, \quad (15)$$

$$\text{S: } q_t^{(s)} = \lambda + \nu p_{t-1} + \eta_t \quad (16)$$

と表現しよう。ただし $\alpha, \beta, \lambda, \nu$ は未知母数、初期値 p_0 を所与とする。このとき需給が均衡すると考えると $q_t = q_t^{(d)} = q_t^{(s)}$ よりクモの巣 (cobweb behavior) 的な変動により p_0 と η_1 により q_1 , 次に q_1 と ϵ_1 により p_1 が決まることになる。価格と数量 p_t, q_t について解くと、価格の変動は ($\beta \neq 0$ のとき)

$$p_t = \frac{\lambda - \alpha}{\beta} + \frac{\nu}{\beta} p_{t-1} + \frac{\eta_t - \epsilon_t}{\beta} \quad (17)$$

により自己回帰過程で表現される。 $|\nu/\beta| < 1$ のとき (もともと対数変換している) とすると需要の弾力性の絶対値が供給の弾力性の絶対値より大きい場合) causal representation を持つ定常解が得られる。

識別条件 (Identification Conditions)

経済学者 Wright (1928) は関税の農産物価格に与える効果をデータから計測する為には、データから需要関数と供給関数を無条件には推定できな

¹⁰ 構造型と誘導型の関係から $\beta - \delta \neq 0$ のとき $\mathbf{Var}(p_t) = [\sigma_\eta^2 + \sigma_\epsilon^2 - 2\sigma_{\epsilon\eta}]/(\beta - \delta)^2$, $\mathbf{Cov}(q_t, p_t) = [\beta\sigma_\eta^2 + \delta\sigma_\epsilon^2 + (-\beta - \delta)\sigma_{\epsilon\eta}]/(\beta - \delta)^2$ が導ける。

いことを認識¹¹、今日の操作変数法と呼ばれる方法の利用を検討した。その後、この問題は識別問題と呼ばれるようになったが、Hsiao (1983) が説明している。パラメトリックモデルの場合には有限次元の母数 θ 、同時分布を $f(\mathbf{y}|\theta)$ ($\theta \in \Theta$) とすると同時分布が同一、すなわち観測上同等な分布 $f(\mathbf{y}|\theta')$ ($\theta' \in \Theta$) のとき $\theta = \theta'$ より母数 θ が一意に定まることを意味する。

同時方程式モデルについては Cowles-Commission の Koopmans-Hood (1953) により数学的には明確に表現された。ここで誤差項の同時分布 $f(\cdot)$ として $(\mathbf{B}, \Gamma, f(\mathbf{u}))$ を構造 S とする。二つの構造 S, S' が観測上等価とは $\mathbf{B}^* = \Theta \mathbf{B}$, $\Gamma^* = \Theta \Gamma$, $f^*(\mathbf{u}_t) = \text{mod}|\Theta|^{-1} f(\Theta^{-1} \mathbf{u}^*)$ を満たす非正則行列 Θ (変換行列と呼ぶ) が存在することと同値である。したがって等価の構造を保持する変換行列 Θ が正則な対角行列のみであることが示されれば、基準化規則 (一般性を失うことなく各方程式の左辺の係数を 1 にとれる) が必要なので母数の識別性が示される。

ある構造方程式群 $\mathbf{A} = (\mathbf{B}, \Gamma)$ を考え、一般性を失うことなく 1 番目の方程式の識別問題を考察してみよう。 $G = G_1 + G_2, K = K_1 + K_2$ として当該方程式に現れる変数の数を G_1, K_1 として行列を分割

$$\mathbf{A} = \begin{pmatrix} \mathbf{B} & \Gamma \end{pmatrix} = \begin{pmatrix} \beta_1 & \mathbf{0} & \gamma_1 & \mathbf{0} \\ \mathbf{B}_1 & \mathbf{B}_2 & \Gamma_1 & \Gamma_2 \end{pmatrix} \begin{matrix} \leftarrow \text{1st row} \\ \leftarrow G-1 \text{ rows} \end{matrix} \quad (18)$$

$\underbrace{\begin{matrix} G_1 & G_2 \end{matrix}}_{G \text{ columns}} \quad \underbrace{\begin{matrix} K_1 & K_2 \end{matrix}}_{K \text{ columns}}$

はそれぞれの要素は $\mathbf{B}, \Gamma$ に対応、

$$\Pi = \begin{pmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{pmatrix} \begin{matrix} G_1 \\ G_2 \end{matrix} \quad (19)$$

$\mathbf{K}_1 \quad \mathbf{K}_2$

とする。このとき線形構造方程式の母係数の識別性について次の結果が成立する。(念のために証明は付論に与えておく。)

定理 1 : (i) この方程式が識別されるための必要十分条件 (rank condition) は構造型母数を表す行列の階数 (rank) について

$$\rho(\mathbf{B}_2, \Gamma_2) = G - 1. \quad (20)$$

¹¹Wright (1915) は H. Moore 氏が書いた景気循環 (business cycles) に関する書籍への書評の中で Moore 氏が推定した pig iron (銑鉄) の需要関数 (例 1 における (10) 式を q_t を p_t に回帰して推定したと思われる) の識別問題を指摘している。

で与えられる。この条件は誘導型母数を表す行列の階数について

$$\rho(\Pi_{12}) = G_1 - 1 \quad (21)$$

と同値である。

(ii) 必要条件は

$$K_2 \geq G_1 - 1 \quad (22)$$

で与えられる。

この条件はある構造方程式の識別には除かれた変数が存在することが一つの必要（ゼロ制約）条件であることを示している。例えば例1の需給モデルでは Wright(1928) が指摘したように需要関数や供給関数の母数を推定することは可能ではない。例えば例1の需要関数では $G = G_1 = 2, K = K_1 = 1, K_2 = 0$ より $K_2 - (2 - 1) < 0$ より識別不能 (under-identified case) である。この場合には伝統的な計量経済学では除かれた変数（例えば魚市場では漁場の天候など需要関数に除外されているシフト変数など）が分析に必要とされた¹²。他方、例2は識別可能な統計モデルである。除外制約の条件は線形ゼロ制約条件と呼ばれているが、識別性には共分散に制約をつける方法などもあり例えば Hsiao (1983) に議論がある。除外制約の場合、 $K_2 = G_1 - 1$ の場合は丁度識別可能 (just-identified)、 $K_2 > G_1 - 1$ の場合は過剰識別 (over-identified) と呼ばれている。同時性 (simultaneity) を持つ構造方程式モデルにおける識別性 (identification) の条件の存在が制限情報最尤法 (LIML, Anderson-Rubin (1949)), 二段階最小二乗法 (TSLS) など計量経済における古典的な推定法が開発されたきっかけであった。

因果連鎖モデル (Causal Chain Model)

Wold (1954) は同時方程式による経済変数のモデリングに対して次のように表現できる因果連鎖 (causal chain) モデル、逐次モデル (recursive sys-

¹² こうしたある構造方程式から除かれた外生変数 (exogenous variables) は操作変数 (instrumental variables) とも呼ばれているが、その利用の妥当性を巡って最近の econometrics においても様々な議論がある。

tems) の利用を提唱した。

$$\mathbf{B} = \begin{pmatrix} \beta_{11} & 0 & 0 & \cdots & 0 \\ \beta_{21} & \beta_{22} & 0 & \cdots & 0 \\ \beta_{31} & \beta_{32} & \beta_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \\ \beta_{G1} & \beta_{G2} & \beta_{G3} & & \beta_{GG} \end{pmatrix}, \quad (22)$$

ただし $|\mathbf{B}| = \prod_{g=1}^G \beta_{gg} \neq 0$ (i.e., $\beta_{gg} \neq 0$ ($g = 1, \dots, G$),) および

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & 0 & 0 & \cdots & 0 \\ 0 & \sigma_{22} & 0 & \cdots & 0 \\ 0 & 0 & \sigma_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & & \sigma_{GG} \end{pmatrix} \quad (23)$$

である。(ただし $\sigma_{gg} \neq 0$ ($g = 1, \dots, G$))。この場合は同時分布はかなり簡単化されるので、統計的分析が容易になる。

ここでの仮定の下では変換行列 $\mathbf{\Theta} = (\theta_{gh})$ を考察することより次の結果が成立するので、Wold の逐次モデルは識別可能 (identified) となる。

定理 2: 仮定 1 の下で次の命題が成り立つ。

- (1) $\mathbf{B}, \mathbf{B}^*$ を三角行列とすると $\mathbf{B}^* = \mathbf{\Theta} \mathbf{B}$ であれば行列 $\mathbf{\Theta}$ は三角行列となる。
- (2) さらに行列 $\mathbf{\Sigma}, \mathbf{\Sigma}^*$ が対角行列であり、行列 $\mathbf{\Sigma}^* = \mathbf{\Theta} \mathbf{\Sigma} \mathbf{\Theta}'$ であれば、 $\mathbf{\Theta}$ は対角行列になる。

証明: (1) 行列 $\mathbf{B}, \mathbf{B}^*$ が三角行列のとき $\Rightarrow$ 任意の分割行列 ($G_1, G_2, G_1 + G_2 = G$) について $\mathbf{\Theta} = \begin{pmatrix} \mathbf{X} & \mathbf{O} \\ \mathbf{Y} & \mathbf{Z} \end{pmatrix}_{G_1}^{G_1}$ となる。

(2) $\mathbf{B}, \mathbf{B}^*$ が三角行列、かつ $\mathbf{\Sigma}, \mathbf{\Sigma}^*$ 対角行列の時 $\mathbf{\Theta} = \begin{pmatrix} \mathbf{X} & \mathbf{O} \\ \mathbf{O} & \mathbf{Z} \end{pmatrix}_{G_1}^{G_1}$ となる。したがって $\mathbf{B}$ 三角行列かつ $\mathbf{\Sigma}$ 対角行列なら $\mathbf{\Theta}$ は対角行列となる。
Q.E.D.

各構造方程式の基準化として $\beta_{gg} = 1$ ($g = 1, \dots, G$) としよう。このと

き構造方程式は

$$\begin{aligned}
 y_{1t} &= -\gamma_{11}z_{1t} - \gamma_{12}z_{2t} - \cdots - \gamma_{1K}z_{Kt} + u_{1t}, \\
 y_{2t} &= -\beta_{21}y_{1t} - \gamma_{21}z_{1t} - \gamma_{22}z_{2t} - \cdots - \gamma_{2K}z_{Kt} + u_{2t}, \\
 &\vdots \\
 y_{Gt} &= -\beta_{G1}y_{1t} - \cdots - \beta_{G,G-1}y_{G-1,t} - \gamma_{G1}z_{1t} - \cdots - \gamma_{GK}z_{Kt} + u_{Gt}
 \end{aligned} \tag{25}$$

と表現される。この場合にはモデルは逐次的であり、最初の方程式により y_{1t} が決まり、2 番目の方程式により y_{2t} が決まる、ということになる。

Wold はこうした逐次モデルを“因果連鎖 (causal chain)” モデルと呼び、詳しく検討している。このタイプのモデルでは誤差項が互いに独立なことを仮定しているので、左辺の内生変数と右辺に説明変数として現れる変数は無相関になり線形回帰モデルの仮定をみたしている。したがって逐次的に最小二乗法を適用することにより構造方程式が推定可能である。

Wold の逐次モデルが妥当であればパス係数を最小二乗法により推定することができるので最小二乗法にもとづくパス解析 (path analysis) が有用となる。

3 構造方程式モデルと DAG

遺伝学者 Wright(1921, 1934) はパス解析 (Path Analysis) を開発した。統計学に関連する研究の流れとしては、DAG(directed acyclic graph) による統計的因果分析 (statistical causal analysis) として Perl(2000, 2009) により様々な精緻化が行われ、疫学、社会学、心理学などでかなり応用されている¹³。グラフ論の記号ではある頂点 (変数) α_i から別の頂点 (変数) α_{i+1} ($i = 1, \dots, n+1$) への矢線が (双方向ではなく) 一方向にのみ存在するとき α_i から α_{i+1} への因果性と解釈される。もし $\alpha_{n+1} = \alpha_1$ (n は正整数) となる頂点が存在しない場合を非循環グラフと呼ばれるが、頂点が有限個の非循環グラフは DAG と呼ばれる。

これまでの議論から応用上では重要と思われる次のような命題が導かれる。

定理 3 : DAG により表現される構造方程式モデルと Wold の causal chain モデルは一致する。

¹³例えば宮川 (2004) が非巡回的グラフの利用を巡る様々な論点を解説している。

例 3：ここで $G = 2$ の簡単な例を挙げておく。

$$(方程式 1) \quad X_1 = \beta_0 + \beta_1 Y_0 + \beta_2 L_1 + \beta_3 L_0 + u_X, \quad (23)$$

$$(方程式 2) \quad Y_2 = \alpha_0 + \alpha_1 X_1 + \alpha_2 Y_0 + \alpha_3 L_1 + \alpha_4 L_0 + u_Y, \quad (24)$$

という構造方程式モデルを例として考えると分かりやすい。 u_X, u_Y はノイズ項、変数は Y_2, X_1, Y_0, L_1, L_0 、行列 $\mathbf{B}$ は下三角行列、ノイズの独立性は仮定しておけば、逐次型 ($G = 2$ であり $\mathbf{B}$ は三角行列) で誤差項の分散共分散行列は対角行列の構造方程式モデルである。この場合には外生変数は L_0, L_1, Y_0 、内生変数は X_1, Y_2 であり DAG で表現可能である。この場合には Path 図を書いてみると直観的に変数間の関係が理解しやすいが、この側面が応用統計で Path 図が利用されている大きな理由と考えられる。

Causal Chain モデルの仮定の下で得られるパス係数 (path coefficients) は因果関係 (causal relations) についての causality measure (因果尺度) と解釈できる。ここで構造型係数 $\mathbf{A} = (\alpha_{gh}) = (\mathbf{B}, \mathbf{\Gamma})$ の各要素を h から g へのパス係数とすると、これは直接効果 (direct effects) を表現していると思えることができる。これに対して誘導型係数 $\mathbf{\Pi} = (\pi_{gh}) = -\mathbf{B}^{-1}\mathbf{\Gamma}$ の各要素を h から g への総合効果を (total effects) を表現していると解釈できる。総合効果と直接効果の差は間接効果 (indirect effects) を意味するが、構造方程式モデルでは内生変数間の関係を通じて間接的にある変数が別の変数を通じて複雑な影響を与え得るのである。

DAG 分析の枠組みに置いてパス係数が因果尺度として十分ではないと考えたのが Pearl(2009) はネットワークモデルの発想から $\mathbf{y}_t (= (y_{jt})(j = 1, \dots, G))$ の同時分布ではなく、外生変数・内生変数の順番を入れ替えて全ての変数 x_i ($i = 1, \dots, p$) として同時分布

$$P(x_1, \dots, x_p) = \prod_{i=1}^p P(x_i | pa_i) \quad (25)$$

という表現から出発している (Pearl (2009), Page 16, (1.33))。ただし $p = G + K$, pa_i は変数 x_i の親・先祖の変数を表す。こうしたネットワークモデルより因果性を定義、Do 作用による介入 (intervention) 効果に基

づく因果尺度を導入している¹⁴。DO 操作とはある変数 $X_i = x_i$ (x_i は固定された数値) で条件付ける際に X_i の親・先祖変数 $pa(x_i)$ を固定して影響を除く操作と解釈しておく。

Causal Chain モデルを仮定すると外生変数を所与としたとき内生変数の同時分布は

$$\begin{aligned} f(y_{1t}, \dots, y_{Gt} | \mathbf{z}_t) &= f(y_{1t} | \mathbf{z}_t) f(y_{2t}, \dots, y_{Gt} | \mathbf{z}_t) \\ &= f(y_{1t} | \mathbf{z}_t) f(y_{2t} | y_{1t}, \mathbf{z}_t) f(y_{3t}, \dots, y_{Gt} | \mathbf{z}_t) \end{aligned} \quad (26)$$

と分解可能である。例えば $pa_2 = \{y_1, \mathbf{z}_t\}$ と定義して (ただし親・先祖でない変数は除く) この操作を繰り返せば、良く知られた同時分布の分解公式のように対応する。こうしたグラフの経路上での条件付した分析が応用上でどのような意味があるか興味深い。例えば econometrics では外生性 (exogeneity) ・内生性 (endogeneity) および政策効果の計測法についての議論が行われている¹⁵ が、変数の条件付分布による特徴づけにより分析が行われている。他方、Pearl (2009, 5 章) は外生性 (exogeneity) は Do 作用素により簡便に定義できることを指摘している。ここで DO 作用素とは DAG モデルに置いて変数 x_i について $pa(x_i)$ を除いた時の分布による評価である。外生性や操作変数を巡る定義は適切な操作変数を実際にいかにして見つけるかという問題と密接に関連しているので重要な意味があると言える。

4 応用例：Noncompliance 問題と操作変数法

例 4 として生物統計分野における重要な一つの統計的問題として non-compliance 問題を取りあげる¹⁶。このモデルは Rubin 流の counterfactual (反実仮想) モデルであるが、変数 Z_{1t} を処理群・対照群への割り付け変数、コンプライアンス変数 D_t (薬を飲むなど実際に割り付けられた処置を実現か否かする)、処置効果変数 $Y_t(D_t)$ とする。 Z_{1t}, D_t のいずれも 1, 0 をとる変数、 $Y_t(D = 1), Y_t(D = 0)$ を服用、非服用の効果を示す変数デとすると、この noncompliance モデルでは無条件では $\mathbf{E}[Y_t(D = 1) - Y_t(D = 0)]$ である ATE (average treatment effects, 平均処理効果) は識別されない。こ

¹⁴Parents (親), Ancenser (先祖) などの定義は例えば宮川 (2004), Pearl (2009) を参照されたい。

¹⁵例えば Engle, Hendry and Richard (1983) が挙げられる。

¹⁶日本語の解説として佐藤 (2006), 岩崎 (2015) などがある。

ここで Z_{1t} , D_t の可能性に 4 つのタイプの被験者 (Complier, Always-taker, Never-taker, Defier, ここでは C, A, N, D で表す。) が存在するとき Z を実験の割り付け変数 (操作変数)、 D を服用変数 (内生変数)、 Y を服用効果変数 (内生変数) とする構造方程式モデルとみなすことが可能であるが、noncompliance モデルでは無条件では ATE (average treatment effects, 平均処理効果) はデータから識別されない。直観的にはある ATE の値をデータで観察できたとしても、C グループと D グループで効果が相殺されるが、相殺の割合を観測データのみから知ることはできないので本当の薬の服用効果は分からないと解釈できる。

Angrist, Imbens and Rubin (AIR, 1996) は処理変数 D_t を割り付け変数 Z_t の関数と見て $D_t(Z_{1t})$ に単調性を仮定した上で操作変数法 (instrumental variables method) の利用を提案している。ここで単調性とは確率変数 $D_t(Z_t)$ が確率変数の単調関数であることを意味するが、Defier は $Z_{1t} = 1, D_t = 0, Z_{1t} = 1, D_t = 0$ なのでこの仮定下では排除される。AIR(1996) の主張は次のように解釈できるが、その証明は付論に与えておく。

定理 4 : (i) Noncompliance モデルは一般に識別可能でない。(観測データから処理効果の差は一意に決まらない。) 処理変数に単調性を仮定すると識別可能となる。

(ii) 単調性の仮定の下で $G = 2$ の逐次型の線形構造方程式モデルに表現される。すなわち、未知母数 $\gamma, \delta, \alpha, \beta$,

$$D_t = \gamma + \delta Z_{1t} + v_t, \quad (27)$$

$$Y_t = \alpha + \beta D_t + u_t \quad (28)$$

と表現される。ただし (i) 母数 γ, δ を適当に選ぶ、(ii) 構造方程式 (28) における誤差項 u_t と変数 D_t は無相関とは限らない、(iii) 誤差項は $\mathbf{E}(v_t) = \mathbf{E}(u_t) = 0$ を満たすが、相関はゼロとは限らない。

単調性の仮定より集合 $Z_{1t} = 1, D_t = 0$ (処理群に割り当てられたので服用しない人々から N を除いた) の中から D (天邪鬼) は排除されている。なお、元々の構造方程式モデルでは変数 D_t は定数と外生変数 Z_{1t} により定義される非線形関数という特徴があり、逐次型の非線形な構造方程式モデルの一種、しかし母数は識別できない解釈できる。そこでこのモデルを線形構造方程式に近似して推定する試みと解釈できる。

この構造方程式モデルは逐次型ではあるが説明変数と誤差項が相関することになり回帰モデルと見なすことはできないので例えば最小二乗推

定を行うと、無視できないバイアスが生じる。ここでの問題設定から変数 Z_{1t} はランダムに割り付けているので外生的 ((28) に現れる u_t とは無相関)、さらに外生変数 Z_{1t} は処理効果を表す構造方程式 (28) には出てこない (exclusion restriction と呼ばれている)。定理 1 の識別の必要十分条件は (27) の $\delta \neq 0$ であり、服用変数 D_t が割付変数 Z_{1t} に依存しているという臨床試験では自然な条件を意味しているが、操作変数の条件 (i),(ii) も意味している。

Angrist, Imbens and Rubin (1996) は単一の構造方程式の推定に操作変数法 (instrumental variables method) の利用を提案している。実は二つの誤差項 v_t, u_t は無相関とは仮定できないので構造方程式の推定では最小二乗法をそのまま適用すると無視できない推定バイアスが生じるのである。

ここで操作変数 (instrumental variables) をより一般的に定義しておこう。(2) 式の一歩目の構造方程式を $y_{1t} = \mathbf{b}' \mathbf{x}_{1t} + u_{1t}$ と表現しよう。ただし $\mathbf{x}_{1t}$ は右辺に現れる内生変数・外生変数で識別性を満たす $p \times 1$ 変数ベクトルとする。次の条件を満たす $p \times 1$ 変数ベクトル $\mathbf{w}_t$ を考えよう。例えば (28) では $p = 2$, $\mathbf{x}_{1t} = (1, D_t)'$, $\mathbf{b} = (\alpha, \beta)'$ である。

- (i) 誤差項 u_{1t} とは漸近的に無相関 $\text{plim}(\frac{1}{T}) \sum_{t=1}^T \mathbf{z}_t u_{1t} = \mathbf{0}$ となる。
- (ii) 右辺の変数ベクトル $\mathbf{x}_{1t}$ とは相関して $(\frac{1}{T}) \sum_{t=1}^T \mathbf{z}_t \mathbf{x}_{1t}' \xrightarrow{p} \mathbf{C}$, ただし $\mathbf{C}$ は正則行列となる。

さらに追加の条件として¹⁷

- (iii) 漸近正規性 $\frac{1}{T} \mathbf{W}' \mathbf{u}_1 \xrightarrow{w} N(\mathbf{0}, \sigma_1^2 \mathbf{A})$ が成立する、ことを用いる。ただし σ_1^2 は誤差項 u_{1t} の分散 (分散均一性を仮定した), $\mathbf{u}_1 = (u_{1t})$ は $T \times 1$ 誤差ベクトルである。

このとき次の命題が成り立つ。

定理 5: 操作変数推定量を

$$\hat{\mathbf{b}} = (\mathbf{W}' \mathbf{X})^{-1} \mathbf{W} \mathbf{y}_1 \quad (29)$$

とする。ただし $n \times 1$ ベクトル $\mathbf{y}_1 = (y_{1t})$ である。 $n \rightarrow +\infty$ のとき

- (a) 条件 (i),(ii) の下で一致性 $\hat{\mathbf{b}} - \mathbf{b} \xrightarrow{p} \mathbf{0}$ を満たす。
- (b) 条件 (i)-(iii) のもとで漸近正規性

$$\sqrt{T}(\hat{\mathbf{b}} - \mathbf{b}) \xrightarrow{w} N(\mathbf{0}, \sigma_1^2 \mathbf{C}^{-1} \mathbf{A} \mathbf{C}^{-1}) \quad (30)$$

¹⁷数理的には Lindeberg 条件などの議論が必要であるが数理統計学・確率論の教科書を参照することとして、ここでは説明を省略する。

となる。

(証明) : (a) 推定量に $\mathbf{y}_1$ を代入すると

$$\hat{\mathbf{b}} - \mathbf{b} = (\mathbf{W}'\mathbf{X})^{-1}\mathbf{W}'\mathbf{u}_1 = \left(\frac{1}{T}\mathbf{W}'\mathbf{X}\right)^{-1}\frac{1}{T}\mathbf{W}'\mathbf{u}_1$$

と表現できる。ここで $\mathbf{u}_1 = (u_t)$ は $T \times 1$ の誤差項ベクトルである。したがって条件 (i),(ii) より一致性が得られる。

(b) 基準化を $\sqrt{T}$ とすると

$$\sqrt{T}(\hat{\mathbf{b}} - \mathbf{b}) = \left(\frac{1}{T}\mathbf{W}'\mathbf{X}\right)^{-1}\frac{1}{\sqrt{T}}\mathbf{W}'\mathbf{u}_1$$

と表現される。ここで右辺に対して条件 (ii), (iii) を利用すると漸近正規性が得られる。 (Q.E.D.)

非線形関係を $D_t = \delta Z_{1t} + v_t$ と近似すると解釈すると、この推定法は二段階最小二乗 (TSLS, two-stage least squares method) を含む興味深い例となっている。ここで二段階最小二乗法 (TSLS, two-stage least squares method) とは第一段階で外生変数を用いて右辺の内生変数を最小二乗法により $\mathbf{z}_t = (1, Z_{1t})'$, $\mathbf{1} = (1, \dots, 1)'$, $\mathbf{D} = (D_t)$ ($T \times 1$) $\mathbf{Z} = (\mathbf{z}_t')$ ($T \times 2$) により操作変数行列

$$\mathbf{W}_2 = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'[\mathbf{1}, \mathbf{D}] \quad (31)$$

を構築する。第二段階ではこの操作変数ベクトルを左辺の内生変数 y_{1t} に回帰、最小二乗法を適用して同時方程式における単一の構造方程式を推定する方法である (制限情報推定法の一つ) と呼ばれている。この場合には $\mathbf{W}_2'\mathbf{X} = \mathbf{X}'\mathbf{P}_Z\mathbf{X}$, $\mathbf{W}_2'\mathbf{W}_2 = \mathbf{X}'\mathbf{P}_Z\mathbf{X}$ (ただし射影 $\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, $\mathbf{X} = [\mathbf{1}, \mathbf{D}]$) となる。ここでとりあげた構造方程式モデル ((27) および (28)) の場合には 2×2 行列 $\mathbf{Z}'\mathbf{X}$ が正則なら $(\mathbf{W}_2'\mathbf{X})^{-1}\mathbf{W}_2'\mathbf{y}_1 = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}_1$ となり二段階最小二乗推定量は操作変数法に一致する¹⁸。さらに漸近分散共分散は $\sigma_1^2\mathbf{C}^{-1}$ と簡単化される。

伝統的な計量経済学では同時方程式モデルにおける識別可能な構造方程式の推定法としては二段階最小二乗法については T.W. Anderson, H. Theil、操作変数法については A. Wald, O. Reiersol, J. Durbin, D. Sargan などの研究が知られているが、漸近的には Anderson-Rubin (1949) の制限情報最尤法 (LIML) を同等になるので分散均一性を仮定するとある種の

¹⁸この場合には LIML 推定量とも一致することを示すことができる。

境界を達成していると見なせる。さらに構造方程式の誤差項同士の相関まで考慮した完全情報推定法としては最尤法 (FIML) , 三段階最小二乗法 (3SLS) などの古典的方法が知られている。また誤差項の分散均一性を仮定できず、分散不均一の場合には EL(empirical likelihood, 経験尤度法), GMM (generalized method of moments, 一般化積率法) が知られているので付論 7-3 に説明しておいた。

近年のデータ分析では個別主体の異質性 (heterogeneity) の議論が盛んである。構造方程式モデルに置いて誤差項が分散不均一な場合には $\hat{u}_t$ ($t = 1, \dots, T$) を残差として利用することにより操作変数推定 (29) の漸近分散 (定理 5(b)) に現れる項 $\sigma_1^2 \mathbf{A}$ を $\text{plim}_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T u_t^2 \mathbf{w}_t \mathbf{w}_t'$ に置き換えると、漸近的にはある種のセミパラメトリック限界 (bound) を達成していることが分かる。また分散が不均一な場合での効率的な推定法も導くことができる。また定理 5 の主張はデータが必ずしも互いに独立でない場合なども含みうることなどは今後の展開の可能性を示唆していると言えよう。

例 4 において構造方程式モデル (28) より係数 β の操作変数推定量

$$\hat{\beta}_{IV} = \frac{\sum_{i=1}^n (Z_{1i} - \bar{Z}_1)(Y_i - \bar{Y})}{\sum_{i=1}^n (Z_{1i} - \bar{Z}_1)(D_i - \bar{D})} \quad (32)$$

の確率極限は単調性の仮定の下で

$$\beta_C = \frac{\text{Cov}(Y_i, Z_{1i})}{\text{Cov}(D_i, Z_{1i})} = \frac{\mathbf{E}[Y_i | Z_{1i} = 1] - \mathbf{E}[Y_i | Z_{1i} = 0]}{\mathbf{E}[D_i | Z_{1i} = 1] - \mathbf{E}[D_i | Z_{1i} = 0]}$$

となる¹⁹ 操作変数の単調性を仮定せず構造方程式が識別できない場合には Defier が存在するので、操作変数推定量の確率極限は

$$\beta_{IV} = \frac{[\mathbf{E}[Y_i | Z_{1i} = 1] - \mathbf{E}[Y_i | Z_{1i} = 0]]_C + [\mathbf{E}[Y_i | Z_{1i} = 1] - \mathbf{E}[Y_i | Z_{1i} = 0]]_D}{[\mathbf{E}[D_i | Z_{1i} = 1] - \mathbf{E}[D_i | Z_{1i} = 0]]_C - \pi_D}$$

と表現される。ここで $[\cdot]_C, [\cdot]_D$ は complier, defier に対応する項、 π_D は Defier の確率である。Always-taker と Never-taker は分母と分子の対応する項

¹⁹例えば $Y_i = Y_i(Z_i, D_i(Z_i))$ において Z_i と D_i は 0 or 1 をとるダミー変数、 $p = P(Z_i = 1)$ ($0 < p < 1$) とする。このとき

$$\begin{aligned} \text{Cov}(Y_i, Z_i) &= \mathbf{E}(Y_i Z_i) - \mathbf{E}(Y_i) \mathbf{E}(Z_i) \\ &= [\mathbf{E}(Y_i Z_i | Z_i = 1)p + \mathbf{E}(Y_i Z_i | Z_i = 0)(1-p)] \\ &\quad - p[\mathbf{E}(Y_i | Z_i = 1)p + \mathbf{E}(Y_i | Z_i = 0)(1-p)] \\ &= p(1-p)[\mathbf{E}(Y_i | Z_i = 1) - \mathbf{E}(Y_i | Z_i = 0)] \end{aligned}$$

となる。(国友 (2025) の (2.6) 式の導出を参照。)

はゼロとなる。ここで complier の確率 π_C を用いて基準化した complier の確率測度を $d P_C = d P / \pi_C$ ($\pi_C > 0$) (P は元の確率測度) とする。Defier (天邪鬼) は complier とは Z_i, D_i について逆の選択をするので、処理効果について $[E[Y_i|Z_{1i} = 1] - E[Y_i|Z_{1i} = 0]]_D = [E[Y_i|Z_{1i} = 0] - E[Y_i|Z_{1i} = 1]]_C$ とすると、 $\beta_{IV} = \beta_C$ となる。すなわち、操作変数推定量は Defier が存在する識別不能なモデルにおいても complier の処理効果の推定量として、ある種のロバストネスを持つということが分かる。念のためであるが以上の考察から得られた結果を次のようにまとめておこう。

系 6： Noncompliance モデルにおいて操作変数推定量 (32) を利用する。
(i) 操作変数の単調性を仮定する。操作変数が Z_{1t} ($t = 1, \dots, T$) のみが利用可能なとき分散不均一性の下で（誤差項の 4 次積率が存在などの正則条件を仮定すると）漸近的に分散の下限を達成している。
(ii) 単調性を仮定せずに C, A, N, D が正の確率で存在することを仮定する。ここで Defier の処理効果が Complier の処理効果の $-d$ ($d > 0$) 倍 $[E[Y_i|Z_{1i} = 1] - E[Y_i|Z_{1i} = 0]]_D = d[E[Y_i|Z_{1i} = 0] - E[Y_i|Z_{1i} = 1]]_C$ となることを仮定する。このとき操作変数推定量の確率極限は $\beta_{IV}(\pi_C - d\pi_D)/(\pi_C - \pi_D)$ で与えられる。特に $d = 1$ なら Complier の処理効果の一致推定量となる。

系 6 において Defier の割合 π_D が小さければ操作変数法の推定バイアスは小さいことを意味しているので、直観的な結果と整合的と言えよう。

なお Imbens-Angrist (1994) は CATE(Complier average treatment effects) を局所平均処置効果 (local average treatment effects, LATE)、あるいは CATE (Compliers' average treatment effects) と呼んでいる。さらに Imbens and Rubin (2015) は One-sided Noncompliance, Two-sided Noncompliance など様々な Noncompliance(非遵守) モデルを考察している。本稿の議論が分析に役に立つと思われるがここでは省略する²⁰。

また関連してこの間はかなり重要な展開としては Manski などによる部分識別法²¹ の応用が挙げられる。識別性がない場合、識別可能でない集合の中で bound(境界) を求める方法を Manski (2003) は部分識別法と呼

²⁰例えば国友 (2025) の例 3, 例 4 の議論を参照されたい。

²¹Partial identification と呼ばれている。識別不能な統計モデルにおいて処理効果の限界 (bound) をノンパラメトリックに求めようとする試みである。例えば奥村 (2018) に解説がある。近年の micro 計量経済では様々な展開がある。

んだが様々な展開がある。ただし、例えばここで問題としている設定では結果変数が有界でない場合になる。また条件付確率の制約からのみでは得られる bound はかなり広くなる。したがって実用的な観点からはなお検討の余地があるが議論は省略する。

5 おわりに：計量経済・計量生物・応用統計の接点

古典的な計量経済学では同時方程式にもとづく構造方程式の推定方法が開発された。推定方法は完全情報推定法と制限情報推定法に分かれるが、いずれも同時性が存在する場合の一般的設定の下での方法である。したがって、構造方程式が逐次的に決定される状況であっても利用可能である。ここでは制限情報推定法として LIML, TSLS, IV、構造方程式の誤差項における従属性を明示的に扱う完全情報推定法として最尤法 (FIML)、三段階最小二乗法 (3SLS) など古典的方法、さらに GMM, EL(経験尤度法) などのより新しい方法についても言及しておいた。こういった計量経済分野で研究されている方法は同時性を含む構造方程式の推定論であるが、狩野 (2002) が説明しているような計量生物や日本の応用統計学で良く利用されている逐次型の構造方程式を含んでいることを強調しておこう。

ただし、逐次型の構造方程式が識別可能であるためには定理 2 の係数および誤差項の共分散についての強い制約が必要である。この条件があるために多くの経済学者には DAG による分析や Pearl (2009) の主張が時には説得的でないと判断する一因と考えられる。構造方程式の誤差項間の相関は臨床試験などの計量生物の設定では無視しても良い場合もあるが、相互関連性の強いことが想定される状況や同時性が重要な市場データでは十分に説得的では無いというのが、ここでの暫定的結論である²²。なお、確率的誤差項の無相関性については識別条件が成り立てば統計的検定を行うことは可能であるが、この問題を含め計量経済学で展開された過剰識別性 (over-identification) 検定や外生性検定などの手法の利用を含めて様々な応用統計の可能性が考えられる。

²²例えば DAG の応用経済学における利用については Imbens (2020) は否定的見解を述べている。米国では G. Imbens(経済学者, 2021 年ノーベル賞 (経済学) 受賞者) と J. Pearl(計算機学者) の間などで因果分析を巡る基本的問題についての論争があるようである。因みに Imbens 氏の論文が掲載されたのはアメリカ経済学会の学術誌であることは興味深い。また Heckman (2008) は経済学的因果関係についての独自の主張を展開している。

最後の話題になるが、計量経済学における重要な分野としてマクロ経済・金融経済の分野では多次元時系列分析の利用が盛んであるが、これまでのところ他分野との関連はあまり議論されていない。

例えば誘導型 (7) における変数ベクトル $\mathbf{z}_t$ を過去の内生変数 $\mathbf{y}_{t-s}$ ($s = 1, \dots, p$) と置けば、多次元自己回帰表現

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{s=1}^p \boldsymbol{\Pi}_s \mathbf{y}_{t-s} + \mathbf{v}_t \quad (t = 1, \dots, T) \quad (33)$$

が得られる。ここで $\boldsymbol{\mu}$ は定ベクトル、 $\boldsymbol{\Pi}_s$ は $G \times G$ ($s = 1, \dots, p$) 係数行列、誤差項の分散・行列共分散 $\boldsymbol{\Omega}$ であるが、データ分析には初期条件 $\mathbf{y}_t$ ($t \leq 0$) が必要となる。例 2 では $G = 2, p = 1$, $\mathbf{y}_t = (q_t, p_t)'$, $\boldsymbol{\mu} = \begin{pmatrix} \lambda \\ (\lambda - \alpha)/\beta \end{pmatrix}$,

$\boldsymbol{\Pi}_1 = \begin{pmatrix} 0 & \gamma \\ 0 & \gamma/\beta \end{pmatrix}$ である。

経済時系列分析の中から多次元時系列データについて Granger (1969) は定常確率過程を仮定した上で時間的な予測可能性による Granger-causality の概念を導入した。時間の要素を統計的因果分析に導入することは重要であるが、議論が複雑化、解決済みというには程遠い段階である²³。この間、構造方程式アプローチを巡る問題を含めマクロ・金融時系列分析では様々な議論が展開されている。こうした問題についての因果分析法についてはなお検討すべき課題がある。

二人の Wright 教授が考察した統計的問題の研究から既に約 100 年が経過した。しかし、二人が提起した問題の正確な理解、計量経済、生物統計、応用統計の基本的問題に関連する問題の検討、などなお課題も少なくない。

²³例えば Kunitomo and Xue (2025) は時系列における Granger Causality (G-Causality) に関連して提案された因果尺度 (G-Causality Measures) の関係をしている。また予想変数や非定常性を巡る基本的問題については例えば国友 (2011) を参照されたい。

参考文献

- [1] Anderson, T.W. (1979), *Statistical Inference in Simultaneous Equations Models*, (Unpublished) Lecture Notes, Stanford University.
- [2] Anderson, T.W. and H. Rubin (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 46-63.
- [3] Anderson, T.W., N. Kunitomo and Y. Matsushita (2011), "On finite sample properties of alternative estimators of coefficients in a structural equation with many instruments," *Journal of Econometrics*, Vol. 165-1, 58-69.
- [4] 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.
- [5] Engle, R., Hendry, D. and J.F. Richard (1983), "Exogeneity," *Econometrica*, Vol. 51-2, 277-304.
- [6] Granger, C. (1969), "Investigating causal relationships by econometric methods and cross-spectral methods," *Econometrica*, 37, 424-438.
- [7] Haavelmo, T. (1943), "The Statistical Implications of a System of Simultaneous Equations," *Econometrica*, Vol.11-1, 1-12.
- [8] Haavelmo, T. (1944), "The probability approach in econometrics," *Econometrica*, Vol.12. Supplement.
- [9] Hayashi, F. (2000), *Econometrics*, Princeton U.P.
- [10] Heckman, J. (2008), "Econometric Causality," NBER Discussion Paper 13934.
- [11] Hsiao, C. (1983), "Identification," in *Handbook of Econometrics*, Vol.1, 223-283
- [12] Imbens, G. (2020), "Potential Outcome and Directed Asyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics," *Journal of Economic Literature*, 58(4): 1129-79.
- [13] Imbens, G. and J. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, Vol. 62, 467-475.

- [14] Imbens, G. and D. Rubin (2015), *Causal Inference for Statistics, Social and Biomedical Sciences*, Cambridge UP.
- [15] Koopmaans, T. and Hood, W. (1953), "The Estimation of Simultaneous Linear Economic Relationships," in *Studies in Econometric Method*, Yale University Press..
- [16] Kunitomo, N. and Xue, Y. (2025), "Akaike's Relative Power Contribution : A Revisit," SSE-DP2025-5, Institute of Statistical Mathematics, <https://stat-expert.ism.ac.jp/wp/wp-content/uploads/2025/06/SSE-DP-2025-5.pdf>
- [17] Manski, C. (2003), *Partial Identification of Probability Distributions*, Springer.
- [18] Pearl, J. (2009), *Causality: Models, Reasoning and Inference*, 2nd Edition, Cambridge UP.
- [19] Owen, A. (2001), *Empirical Likelihood*, Chapman-Hall.
- [20] Qin, J. and Lawless, J. (1994), "Empirical likelihood and general estimating equations," Vol.22, 300-325, *The Annals of Statistics*.
- [21] Wold, H. (1954),"Causality and Econometrics," *Econometrica*, 22-2, 162-177.
- [22] Wright, P. (1915),"Moore's Economic Cycles," Vol. 29-3, 631-641, *The Quarterly Journal of Economics*.
- [23] Wright, S. (1921),"Correlation and Causation," XX-7, 558-585, *Journal of Agricultural Research*.
- [24] Wright, P. (1928), *The Tariff of Animal and Vegetable Oils: Appendix B*, The MacMillan Company. Reproduced in *The Econometrics Journal*, 28(1):120, (2025).
- [25] Wright, S. (1934),"The Method of Path Coefficients," 161-215, *The Annals of Mathematical Statistics*.
- [26] 池田真也 (2025)," 操作変数法の起源 : 二人の Wright 教授と IV ," 未定稿.
- [27] 岩崎学 (2015), 統計的因果推論, 朝倉書店.
- [28] 奥村綱雄 (2018), 部分識別入門, 日本評論社.

- [29] 国友直人 (2011), 構造方程式モデルと計量経済学, 朝倉書店.
- [30] 国友直人 (2025), ”操作変数法の理解へ: 計量生物と計量経済の邂逅,” 日本統計学会誌, 54-2, 109-143.
- [31] 狩野裕 (2002), 「構造方程式モデリング, 因果推論, そして非線形性」, 岩波書店. 統計科学のフロンティア 5, 多変量解析の展開.
- [32] 佐藤俊哉 (2006), 「ランダム化に基づく解析」, 朝倉書店. 臨床試験ハンドブック 24 章.
- [33] 宮川雅己 (2004), 統計的因果推論, 宮川雅己, 朝倉書店.

数理的補論

この付論では本文で示した定理の証明および補足的説明を与えるが、定理 1 の証明は Anderson (1979) の講義録における証明の日本語版である。

(7.1) 定理 1 の証明

(i) 変換行列 Θ を分割して

$$\Theta = \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \Theta_{22} \end{pmatrix} \begin{array}{l} \leftarrow \text{1st row} \\ \leftarrow G-1 \text{ rows} \end{array} \quad (\text{A.34})$$

1 column $G-1$ columns

としよう。この行列を左から $\mathbf{A}$ に乗じると

$$\begin{aligned} \mathbf{A}^* = \Theta \mathbf{A} &= \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \Theta_{22} \end{pmatrix} \begin{pmatrix} \beta_1 & \mathbf{0} & \gamma_1 & \mathbf{0} \\ \mathbf{B}_1 & \mathbf{B}_2 & \Gamma_1 & \Gamma_2 \end{pmatrix} \\ &= \begin{pmatrix} \theta_{11}\beta_1 + \theta_{12}\mathbf{B}_1 & \theta_{12}\mathbf{B}_2 & \theta_{11}\gamma_1 + \theta_{12}\Gamma_1 & \theta_{12}\Gamma_2 \\ \theta_{21}\beta_1 + \Theta_{22}\mathbf{B}_1 & \Theta_{22}\mathbf{B}_2 & \theta_{21}\gamma_1 + \Theta_{22}\Gamma_1 & \Theta_{22}\Gamma_2 \end{pmatrix} \end{aligned} \quad (\text{A.35})$$

G_1 columns G_2 columns K_* columns K_{**} columns

となる。ここで二つの構造が $\mathbf{A}^* \sim \mathbf{A}$ (等価) となる条件は

$$\mathbf{A}^* = \begin{pmatrix} \beta_1^* & \mathbf{0} & \gamma_1^* & \mathbf{0} \\ \mathbf{B}_1^* & \mathbf{B}_2^* & \Gamma_1^* & \Gamma_2^* \end{pmatrix}, \quad (\text{A.36})$$

より

$$\underset{1 \times (G-1) \quad (G-1) \times (G_2+K_2)}{\boldsymbol{\theta}_{12}(\mathbf{B}_2, \boldsymbol{\Gamma}_2)} = \mathbf{0}. \quad (\text{A.37})$$

で与えられる。したがって変換行列 $\boldsymbol{\Theta}$ の最初の行が (1,1) 要素を除いてゼロになる条件 $\boldsymbol{\theta}_{12} = \mathbf{0}$ より第 1 行の識別性の階数条件 (*rank condition*)

$$\rho(\mathbf{B}_2, \boldsymbol{\Gamma}_2) = G - 1 \quad (\text{A.38})$$

が得られる。

(ii) 誘導型母数

$$\boldsymbol{\Gamma} = -\mathbf{B}\boldsymbol{\Pi} \quad (\text{A.39})$$

を書き換えると

$$\begin{aligned} \begin{pmatrix} \gamma_1 & \mathbf{0} \\ \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2 \end{pmatrix} &= - \begin{pmatrix} \boldsymbol{\beta}_1 & \mathbf{0} \\ \mathbf{B}_1 & \mathbf{B}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\Pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\Pi}_{21} & \boldsymbol{\Pi}_{22} \end{pmatrix} \\ &= - \begin{pmatrix} \boldsymbol{\beta}_1 \boldsymbol{\Pi}_{11} & \boldsymbol{\beta}_1 \boldsymbol{\Pi}_{12} \\ \mathbf{B}_1 \boldsymbol{\Pi}_{11} + \mathbf{B}_2 \boldsymbol{\Pi}_{21} & \mathbf{B}_1 \boldsymbol{\Pi}_{12} + \mathbf{B}_2 \boldsymbol{\Pi}_{22} \end{pmatrix}. \end{aligned} \quad (\text{A.40})$$

が得られる。したがって

$$\boldsymbol{\beta}_1 \boldsymbol{\Pi}_{12} = \mathbf{0}, \quad \gamma_1 = \boldsymbol{\beta}_1 \boldsymbol{\Pi}_{11} \quad (\text{A.41})$$

となる。ここで母数ベクトル $\boldsymbol{\beta}_1, \gamma_1$ が定数倍を除いて一意に決まるには

$$\rho(\boldsymbol{\Pi}_{12}) = G_1 - 1 \quad (\text{A.42})$$

が必要十分である。(条件下で最初の方方程式を $\boldsymbol{\beta}_1$ について解き、次に γ_1 について解けばよい。) この階数条件より必要条件

$$K_2 \geq G_1 - 1 \quad (\text{A.43})$$

が得られる。

(iii) これまで識別の必要十分条件として二つの条件を導いた。これらの条件を A, Π とすると等価であることを示す。

補題 A-1. 階数 (rank) $\rho(\mathbf{B}) = G$ であれば $\rho(\mathbf{B}_2) = G_2$.

証明: 行列 $\mathbf{B}$ は正則となるので $\mathbf{B}$ の各列は線形独立となる。したがっ

て行列 $\begin{pmatrix} 0 \\ \mathbf{B}_2 \end{pmatrix}$ の各列は線形独立となるので $\mathbf{B}_2$ の各列も線形独立となる。
Q.E.D.

補題 A-2: $\rho(\mathbf{B}_2, \mathbf{\Gamma}_2) = \mathbf{G} - 1 \Leftrightarrow \rho(\mathbf{B}_2, \mathbf{B}_1 \mathbf{\Pi}_{12}) = \mathbf{G} - 1$.

証明: 階数条件 $\rho(\mathbf{B}_2, \mathbf{\Gamma}_2) = \mathbf{G} - 1$ は行列 $(\mathbf{B}_2, \mathbf{\Gamma}_2)$ の各行が線形独立と同一である。このときある (列) ベクトルが存在して $\mathbf{w}((\mathbf{G} - 1) \times 1)$

$$\mathbf{w}'(\mathbf{B}_2, \mathbf{\Gamma}_2) = \mathbf{0} \Rightarrow \mathbf{w}' = \mathbf{0}. \quad (\text{A.44})$$

となる。これを $\mathbf{\Gamma}_2 (-(\beta_1 \mathbf{\Pi}_{12} + \mathbf{B}_2 \mathbf{\Pi}_{22}))$ に代入すると

$$\mathbf{w}'(\mathbf{B}_2 \mathbf{B}_1 \mathbf{\Pi}_{12} + \mathbf{B}_2 \mathbf{\Pi}_{22}) = \mathbf{0} \Rightarrow \mathbf{w}' = \mathbf{0}, \quad (\text{A.45})$$

と書ける。このとき $\mathbf{w}'\mathbf{B}_2 = \mathbf{0}$, $\mathbf{w}'(\mathbf{B}_2 \mathbf{B}_1 \mathbf{\Pi}_{12}) = \mathbf{0} \Rightarrow \mathbf{w}' = \mathbf{0}$, となり、これは $\rho(\mathbf{B}_2, \mathbf{B}_1 \mathbf{\Pi}_{12}) = \mathbf{G} - 1$ を意味する。 Q.E.D.

補題 A-3: $\rho(\mathbf{EF}) \leq \min[\rho(\mathbf{E}), \rho(\mathbf{F})]$.

以上の補題を利用すると次の結果が得られる。

定理 A-1 : 条件 A $\Leftrightarrow$ 条件 II.

証明: (I) 条件 A $\Rightarrow$ 条件 II.

条件 A より $\rho(\mathbf{B}_2 \mathbf{B}_1 \mathbf{\Pi}_{12}) = \mathbf{G} - 1$ (補題 A-2) となる。したがって $\rho(\mathbf{B}_1 \mathbf{\Pi}_{12}) \geq \mathbf{G}_1 - 1$ であるが、これは $\rho(\mathbf{\Pi}_{12}) \geq \mathbf{G}_1 - 1$ (補題 A-3) を意味する。ここで (トリビアルでないベクトル β_1 に対して) $\beta_1 \mathbf{\Pi}_{12} = \mathbf{0}$ ($\mathbf{B}$ は正則なので最初の行は $\mathbf{0}$ ではない), となり $\rho(\mathbf{\Pi}_{12}) \leq \mathbf{G} - 1$. したがって $\rho(\mathbf{\Pi}_{12}) = \mathbf{G} - 1$.

(II) 条件 II $\Rightarrow$ 条件 A.

条件 $\mathbf{w}'(\mathbf{B}_2 \mathbf{B}_1 \mathbf{\Pi}_{12}) = \begin{pmatrix} 0 & 0 \end{pmatrix}$ を満たすベクトル $\mathbf{w}$ を考える。すなわち $\mathbf{w}'\mathbf{B}_2 = \mathbf{0}$, $\mathbf{w}'\mathbf{B}_2 \mathbf{\Pi}_{12} = \mathbf{0}$ より条件 II および $\beta_1 \mathbf{\Pi}_{12} = \mathbf{0}$ から $\mathbf{w}'\mathbf{B}_2$ は定数倍の β_1 、つまり $\mathbf{w}'\mathbf{B}_1 = k\beta_1$ となるベクトルが存在する。したがって

$$\begin{aligned} (-k \ \mathbf{w}')\mathbf{B} &= (-k \ \mathbf{w}') \begin{pmatrix} \beta_1 & \mathbf{0} \\ \mathbf{B}_1 & \mathbf{B}_2 \end{pmatrix} \\ &= \begin{pmatrix} -k\beta_1 + \mathbf{w}'\mathbf{B}_1 & \mathbf{w}'\mathbf{B}_2 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \end{pmatrix} \end{aligned}$$

となる。ここで $\mathbf{B}$ は正則なので $-k\mathbf{w}' = \mathbf{0}$ 、これは $\mathbf{w}' = \mathbf{0}$ を意味する。したがって $\mathbf{w}'(\mathbf{B}_2 - \mathbf{B}_1\Pi_{12})$ なら $\mathbf{w}' = \mathbf{0}$ となる。これは $(\mathbf{B}_2, \mathbf{B}_1\Pi_{12})$ がフル・ランクとなることを意味するので条件 A が成り立つ。 Q.E.D.

(7.2) 定理 4 の証明

(i) 観測データから推定可能な確率 $p_{11} = P(D = 1, Z = 1), p_{10} = P(D = 1, Z = 0), p_{01} = P(D = 0, Z = 1), p_{00} = P(D = 0, Z = 0)$ とする。(確率の基準化すると 4-1=3 個の母数がある。) 治験参加者には, C, A, N, D の 4 種類であるが、添字を利用して $p_{11} = \pi_{11.C} + \pi_{11.A}, p_{10} = \pi_{10.A} + \pi_{10.D}, p_{01} = \pi_{01.N} + \pi_{10.D}, p_{00} = \pi_{00.N} + \pi_{00.C}$ と表現すると。 $\pi_{11.C}, \pi_{11.A}, \pi_{10.A}, \pi_{10.D}, \pi_{01.N}, \pi_{10.D}, \pi_{00.N}, \pi_{00.C}$ は確率の基準化より 8-1=7 個の母数がある。

ここで平均処理効果は

$$\begin{aligned} & \mathbf{E}[Y(D = 1) - Y(D = 0)] \\ &= [\mathbf{E}[Y(D = 1)|Z = 1]P(Z = 1) + \mathbf{E}[Y(D = 1)|Z = 0]P(Z = 0)] \\ & \quad - [\mathbf{E}[Y(D = 0)|Z = 1]P(Z = 1) + \mathbf{E}[Y(D = 0)|Z = 0]P(Z = 0)] \\ &= [\mathbf{E}[Y(D = 1)|Z = 1] - \mathbf{E}[Y(D = 0)|Z = 1]P(Z = 1)] \\ & \quad + [\mathbf{E}[Y(D = 1)|Z = 0] - \mathbf{E}[Y(D = 0)|Z = 0]P(Z = 0)] \end{aligned}$$

と分解できる。この期待値の各項は条件付確率 $P(D = 1, Z = 1|Z = 1), P(D = 1, Z = 0|Z = 0), P(D = 0, Z = 1|Z = 1), P(D = 0, Z = 1|Z = 0)$ により一意に決まるが、条件付確率は確率 $\pi_{11.C}, \pi_{11.A}, \pi_{10.A}, \pi_{10.D}, \pi_{01.N}, \pi_{10.D}, \pi_{00.C}$ の関数、したがって元の母数は一意に対応しない。したがって無条件では識別可能ではない。

(ii) ここで $P(Z = 1) = p_1, P(Z = 0) = p_0$ 外生的に与えられる変数であることに注目する。さらに割り当てられた被験者は薬を服用できないとすると $\pi_{10.D} = 0$ である。変数 $D_t(Z_{1t})$ の単調性を仮定を利用すると $\pi_{01.D} = 0$ である。このとき $P(D = 1|Z = 1) = (\pi_{11.C} + \pi_{11.A})/p_1, P(D = 1|Z = 0) = \pi_{10.AC}/p_0, P(D = 0|Z = 1) = \pi_{01.A}/p_1, P(D = 0|Z = 0) = (\pi_{00.N} + \pi_{00.C})/p_0$, であるから確率の基準化より 7-2-2=3 個の母数で表現される。

(iii) 方程式 $D = \alpha + \beta Z + v$ と表現して D, Z を代入すると、確率変数 v を $1 = \alpha + \beta + v$ ($D = 1, Z = 1$), $1 = \alpha + v$ ($D = 1, Z = 0$), $0 = \alpha + \beta + v$ ($D = 0, Z = 1$), $0 = \alpha + v$ ($D = 0, Z = 0$) を満たす必要が

ある。したがって

$$\begin{aligned}\mathbf{E}[v] &= (1 - \alpha - \beta)p_{11} + (1 - \alpha)p_{10} + (-\alpha - \beta)p_{01} + (-\alpha)p_{00} \\ &= (p_{11} + p_{10}) - \alpha(p_{11} + p_{10} + p_{01} + p_{00}) - \beta(p_{11} + p_{01})\end{aligned}$$

となる。これより $p_{1\cdot} = p_{11} + p_{10}, p_{\cdot 1} = p_{11} + p_{01}$ とおくと $\mathbf{E}(v) = 0$ となる。

(iv) Noncompliance モデルにおける説明変数 D_t と誤差項 u_t の相関については、国友 (2025) 付論 A で説明しているのでここでは省略する。

(7.3) 経験尤度法と一般化積率法 :

攪乱項の分散が不均一な場合についてデータが互いに独立な確率変数の実現値の場合、ノンパラメトリック尤度関数を考察する。ベクトル列 $\mathbf{y}_t$ ($t = 1, \dots, T$) がある確率分布 $F(\mathbf{y})$ から独立な標本, $P(\mathbf{y}_t = \mathbf{y}) = p_t$ ($t = 1, \dots, T$) のときノンパラメトリック尤度関数は

$$L(F) = \prod_{t=1}^T p_t \quad (\text{A.46})$$

と表現できる。この尤度関数を確率としての制約条件 $p_j \geq 0, \sum_{t=1}^T p_t = 1$ の下で最大化すると解は $p_t = 1/T$ ($t = 1, \dots, T$) となる。

構造方程式モデルでは

$$\sum_{j=1}^T p_t g(y_t, \mathbf{x}_t, \mathbf{w}_t | \mathbf{b}) = 0 \quad (\text{A.47})$$

という制約条件が得られる。(ここで $g(\cdot)$ は構造方程式を意味するが、母集団では $\mathbf{E}[\mathbf{w}_t(y_t - \mathbf{b}'\mathbf{x}_t)] = 0$ 、一般には非線形条件でもよい。 $\mathbf{x}_t$ は $p \times 1$ 、操作変数 $\mathbf{w}_t$ は $K \times 1$ 、 $p = K$ としておく。(なおより一般に $K > p$ の場合は過剰識別となるので議論を修正する必要がある。)

この制約条件および確率 $p_t \geq 0, \sum_{t=1}^T p_t = 1$ の下でノンパラメトリック尤度関数を最大化して母数 $\mathbf{b}$ を推定する方法を経験尤度 (empirical likelihood) 推定と呼ぶと、国友 (2011)3.2 節の議論を少し変更すると

$$\begin{aligned}& \left[\sum_{t=1}^T \hat{p}_t \mathbf{x}_t \mathbf{w}_t' \right] \left[\sum_{t=1}^T \hat{p}_t u_t(\hat{\mathbf{b}})^2 \mathbf{w}_t \mathbf{w}_t' \right]^{-1} \left[\frac{1}{T} \sum_{t=1}^T \mathbf{w}_t y_{1t} \right] \\ &= \left[\sum_{t=1}^T \hat{p}_t \mathbf{x}_t \mathbf{w}_t' \right] \left[\sum_{t=1}^T \hat{p}_t u_t(\hat{\mathbf{b}})^2 \mathbf{w}_t \mathbf{w}_t' \right]^{-1} \left[\frac{1}{T} \sum_{t=1}^T \mathbf{w}_t \mathbf{x}_t' \right] \hat{\mathbf{b}}_{EL}\end{aligned} \quad (\text{A.48})$$

と表現される。ここで $u_t(\hat{\mathbf{b}}) = y_{1t} - \hat{\mathbf{b}}' \mathbf{x}_t$ なので、GMM (generalized method of moments) 推定量は $p_t = 1/T$ ($t = 1, \dots, T$) のとき得られる。したがって MEL 推定量は GMM 推定の一般化となっていることが分かる。

特に $p = K$ のとき操作変数が退化せず条件 $\text{rank}((1/T) \sum_{t=1}^T \mathbf{w}_t \mathbf{x}_t') = p$ を満たしていれば、上記の経験尤度推定量 $\hat{\mathbf{b}}_{EL}$ は操作変数推定量 $\hat{\mathbf{b}}_{IV} = (\sum_{t=1}^T \mathbf{w}_t \mathbf{x}_t')^{-1} (\sum_{t=1}^T \mathbf{w}_t y_{1t})$ に一致する。

なお一般的な非線形の場合について Qin & Lawless (1994), Owen (2001) は漸近的な経験尤度推定量の性質を調べ、観測データ数 n が大きくなる漸近的意味でかなり一般的な正則条件の下で一致性、漸近的正規性、漸近的有效性があることを示している。例えば線形構造方程式モデルの場合には $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ の漸近共分散行列は

$$\mathbf{AEL}(\boldsymbol{\theta}) = [\text{plim}_{n \rightarrow \infty} (\frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{w}_t')]^{-1} [\text{plim}_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2 \mathbf{w}_t \mathbf{w}_t'] [\text{plim}_{T \rightarrow \infty} (\frac{1}{T} \sum_{t=1}^T \mathbf{w}_t \mathbf{x}_t')]^{-1} \quad (\text{A.49})$$

で与えられる。誤差項が均一分散である場合には $\text{plim}_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2 \mathbf{w}_t \mathbf{w}_t' = \sigma_1^2 [\text{plim}_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbf{w}_t \mathbf{w}_t']$ より定理 5 で与えられた漸近的分散共分散行列 (30) に一致する。ただし、データ数が有限の場合には必ずしも EL, GMM が TSLS, LIML よりも良い結果をもたらすとも限らないことを Anderson-Kunitomo-Matsushita (2011) が例示している。なお $K > p$ の場合には (48) を

$$\{[\text{plim}_{n \rightarrow \infty} (\frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{w}_t')] [\text{plim}_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2 \mathbf{w}_t \mathbf{w}_t']^{-1} [\text{plim}_{T \rightarrow \infty} (\frac{1}{T} \sum_{t=1}^T \mathbf{w}_t \mathbf{x}_t')]\}^{-1}$$

に修正する必要がある。また、経験尤度推定量の漸近分布は古典的な状況では GMM 推定法における効率的 GMM 推定量と同等である。計量経済学における GMM 法に関する利用については Hayashi (2000) が詳しい。

おわりに

統計数理研究所が推進している統計エキスパート養成プログラムでは必ずしも統計学を専門としているわけではない各分野の若手研究者とメンターにより統計エキスパート演習を行っているが、直接の関係者以外はこの統計エキスパート養成プログラムの中で具体的にどのような講義や議論が行われているかはよく知られていないのではないかとと思われる。このプロジェクトは統計数理研究所が文部科学省の支援を受け、統計人材育成コンソーシアム (<https://stat-expert.ism.ac.jp/>) における事業の一環として統計エキスパート養成計画を2021年より実施している。統計エキスパート人材育成の事業内容は多岐に及び、この報告書はそのごく一部分に過ぎないが、エキスパート研修の重要部分である全体研修の議論から発生した一つの話題についての共同研究であり、検討した議論をここで例示しておくことも統計エキスパート事業での教育的意味があるのではないかと判断している。

著者の池田氏は開発経済・農業経済を専門とする経済学者、国友は経済・経営・金融などにおける統計的方法に関心のある統計学メンターである。なお、著者は二人とも今回取り上げた話題について普段から専門的に研究していたわけではない。したがって話題について関心があり研究を重ねている専門家から見ると不適切な説明や誤りを含めて様々なコメントがあると思われる。

日本における統計科学分野における統計エキスパート人材の必要性が叫ばれている中で日本における統計科学および応用の諸分野における統計エキスパート養成の教育として妥当な内容は何か、今でもなお模索中である。今後の教育内容を改善する参考の為に、各界の統計科学や応用諸分野などの関係者からの忌憚のないコメントを期待したい。

国友直人 (著者代表)

2025年10月