

LATE vs. 構造推定 — 政策科学の最先端で何が起こっているのか —

竹田 聡

LATE versus Structural Estimation: Considering the Credibility Revolution

Satoshi Takeda

はじめに

社会科学において、政策科学の実証分析手法の刷新を迫る「信頼性革命 (Credibility Revolution)」と呼ばれる学問的な「革命」が起きている。これが露わになったのは2010年のことである。すなわち、米国経済学会の学術誌である、*Journal of Economic Perspectives* 誌 (第24巻第2号) と *Journal of Economic Literature* 誌 (第48巻第2号) が2010年に実証分析手法を巡る特集号を公刊したのである。

この「革命」について、我が国で広く一般向けに初めて紹介したのは、筆者の知る限り、森田 (2014) であると思う¹⁾。森田 (2014) では、「LATEと構造推定の仁義なき戦い」²⁾が考察されている。

本稿では、「信頼性革命」を巡る複数の文献を紹介しながら、EBPM (Evidence-Based Policy Making; エビデンスに基づく政策形成) を巡る諸問題を考察し、地域政策学の将来像を考える上での基礎的かつ重要な知見を整理する。

1. 構造推定とは

構造推定 (structural estimation) とは、経済主

体の意思決定や政策が効果をもたらす構造について、一般的には経済学の理論に基づいて数理モデルを考案し、当該構造モデルの様々なパラメータを推定していく手法である。パラメータが推計されれば、そのモデルを用いて政策シミュレーションを行うことができる³⁾。

元来、経済学の理論は観測された社会現象やデータを基に考案されたものであろうが、その経済学の理論が現実世界の真の構造を正確に記述したものであるとは限らない。

この意味で、構造推定は理論に基づいているが、仮説に過ぎない理論を出発点にして構築されており、その理論仮説の妥当性は必ずしも検証されていない。つまり、構造推定において、理論仮説の一部は前提として仮定されている。

構造推定には、前提として仮定された理論仮説の一部に「欺瞞 (con)」があるとして、実証研究からこれを取り除くべきだと主張したのがヨシュア・アングリストとヨーン・シュテファン・ピスケの2010年論文である (Angrist and Pischke 2010)。「実証経済学における信頼性革命」と題するその論文のサブタイトルは、「リサーチデザインの改善によって計量経済学からどのようにして欺瞞を取り除くの

- 1) 本稿第2節で後述するLATEの旗手であるアングリストとピスケによる計量経済学のテキストブック (Angrist and Pischke 2009) の日本語訳が出版されたのは2013年である。しかし、それよりも森田 (2014)の方が我が国の社会学者集団に対する影響力が圧倒的に強かったように思われるし、そもそも森田 (2014)の基になった森田果氏の『法学セミナー』(第681-707号, 日本評論社)での連載「法律家のための実証分析入門」はもっと早く公刊されている。
- 2) 森田 (2014) 240頁。森田 (2014) は最近の社会科学の実証研究の潮流を知る上でとても有益であると同時に、統計学の初学者でも米国のトップジャーナルに登場する最先端の実証研究を読むことができる、素晴らしい本である。もともと『法学セミナー』に連載されていたものに加筆修正した本なので、統計学を駆使した実証研究の作法を知らない法学徒や法律の実務家向けに書かれたものである。本稿の執筆においても大いに参考にした。
- 3) この文章の「経済学」を「政治学」に、「経済主体」を「各アクター」と読み替えていただいても構わない。言い換えると、「信頼性革命」を巡る議論は、社会科学の実証分析全般に妥当するだろう。

か」という刺激的なものである。

これは実証研究における分析手法を、「欺瞞」に満ちた構造推定ではなく、後述するLATEに刷新しているという「革命」を宣言するものであった。

2. LATEとは

LATEとは

LATEとは、Local Average Treatment Effect (局所的平均処置効果。LATEは「レイトウ」と読む)の略称である。Treatment (処置)とは、社会科学の世界では政策的介入を指す。LATEは政策的介入の因果効果を統計的に分析する手法であり、多少の誤解を恐れず端的に言えば、新薬の治験と同じ手法である。

すなわち、ある集団(これを処置群と呼ぶ。新薬の治験の場合は、新薬を投与された人たち)に対して政策(新薬の投与)を実施(施策)したときの成果を、属性を同じにした別の集団(これを対象群とか、コントロール群と呼ぶ。新薬の治験の場合は、^{プラシーボ}偽薬を投与された人たち)の成果と比較して、統計的に有意な差があるかどうかを検証する手法である。この成果の差を、政策の因果効果と呼ぶ⁴⁾。

処置群と対照群の属性を同じにする、という点がLATEの肝となる。学問は究極的には分類と比較であるとするならば、実証研究において統計学を用いた比較は極めて重要な分析手法となる。統計学を用いた比較において重要なことは、被験者の属性といった、目的変数(被説明変数)に影響を与える処置の有無以外の他の要因を揃えて、局所的平均処置効果を比較するという点である。

他の要因を揃えて比較する

因果関係を考える際に、よく紹介される例として、小中学生の学力と朝食の有無がある。すなわち、朝食を食べる小中学生の方がそうでない小中学生よりも学力が高いという調査結果が出たとする。そこで、学力を目的変数とし、朝食の有無と親の年収を説明変数として回帰分析を行うと、朝食の有無と親の年収が統計的に有意な結果になったとする。この事実から朝食を食べることは学力を向上させる、という因果関係を主張することは誤りである。

なぜか。学力にも朝食の有無にも影響を与える第3の要因を無視しているからである。この第3の要因は、交絡変数(confounding variable)とか、交絡因子(confounding factor, confounder)とか、潜伏変数(lurking variable)などと呼ばれる。この場合、交絡変数は家庭の教育熱心さであり、教育熱心な家庭は子供に朝食を食べさせていて、教育熱心であるがゆえに子供の学力が高い。つまり、学力と朝食の有無の関係は因果関係ではなく、擬似相関(spurious correlation)に過ぎない。因果関係を正しく検証するためには、交絡変数をコントロールする必要がある⁵⁾。

3. 処置群と対照群をどうつくるか

自己選択バイアスとは

他の要因を揃えて比較する、つまり属性を同じにする処置群と対照群をどうやってつくるのか、これがLATEの肝となる。

たとえば、いわゆる「就職氷河期世代」と呼ばれる2018年次点で30代半ばから40代前半までの人たちの一部に、政府が職業訓練等の就職支援プログラム

4) 本稿第6節で後述するエビデンスの階層に鑑みれば、LATEの説明はエビデンスの階層ごとにもっと詳細にすべきであるが、ここではわかりやすさを優先して大胆に簡易的な説明をした。また、新薬の治験の場合は、二重盲検(double blind)といって、新薬や偽薬を投与された人たちだけでなく、投与した医師や看護師にも偽薬であるかどうかなどを知らせずに治験が行われることが一般的である。なお、処置効果をLATEのように検証するのは、薬学や疫学(epidemiology)ではもはや常識であるだろう。

5) この事例で説明変数に親の年収を追加したのは、朝食の有無が子供の学力に与える因果効果を考察するために、目的変数(この場合は、子供の学力)に影響を与える親の年収をコントロールするためである(親の年収が高い家庭の方が塾通いなどの教育支出を行うことができる)。回帰分析において、目的変数に影響を与えそうな要因を「コントロール変数」と称して説明変数に追加することはよく行われる。これは欠落変数バイアス(omitted variable bias)を避けるためである。

を行い、プログラムの受講者の所得を引き上げようとする政策を行ったとする⁶⁾。

このプログラムの政策の因果効果を測定する場合、プログラムを受講した人たちを処置群、プログラムを受講しなかった人たちを対象群として、両者の所得の平均値を比較したとする。仮に処置群の年収が対照群の年収より50万円高かったとして、この50万円の年収の増加を、このプログラムの政策の効果と見てよいだろうか。

結論から言えば、こうした比較は誤りである。というのは、プログラムの受講者（処置群）は、自発的にそうしたプログラムを受講して自分の職業能力を高めようとする点で意欲的な人たちであり、そうでない人たち（対象群）よりも性格スキルも含めた職業能力が元々高い人が多いからだ。処置群には対照群よりも、元々の職業能力が高い人が集まってしまったために、処置群の年収が高くなってしまふ。これは自己選択バイアス（self-selection bias）と呼ばれ、選択バイアスの一つに数えられる。これは、処置群と対照群の属性を系統的（systematic）に異なるものにしてしまふ。

言い換えると、自己選択バイアスが存在する場合、就職支援プログラムという処置以外の他の要因が、目的変数（この場合は、年収）に影響を与えてしまふ。

自己選択バイアスを回避して、政策の因果効果を分析するには、どうすればよいのだろうか。

操作変数法の斬新奇抜な活用

こうしたバイアスを回避するため、操作変数法（instrumental variable approach）という古くからある手法を、斬新奇抜なやり方で活用したのが Angrist (1990) である。Angrist (1990) は、米国において、ベトナム戦争のときの従軍経験という処置が人々の賃金を下げるといふ因果効果を実証分析したものだ。

仮に単純に従軍した人たちを処置群、しなかった

人たちを対象群として、両者の賃金の平均値を比較したとする。処置群の平均賃金が対照群の平均賃金よりも低かったとしても、元々賃金の稼得能力が低い人が従軍するということかもしれない。つまり、自発的に従軍する場合、自己選択バイアスが生じてしまふ蓋然性が高い。

ベトナム戦争のときの米国では、一時期、クジによる徴兵制が行われていたため、Angrist (1990) は、クジの当選の有無を操作変数（IV; Instrumental Variable）とした。クジの当選の有無はランダムで（規則性がないので）、直接的には目的変数（この場合は、賃金）に影響を与えないからだ。

Angrist (1990) は、第1段階として、従軍の有無（従軍確率）を目的変数（被説明変数）とし、クジの当選の有無を説明変数として回帰分析を行い、第2段階として、賃金を目的変数とし、第1段階で推計した従軍確率の推計値を説明変数として回帰分析を行う二段階最小二乗法を行った。その結果、従軍経験という処置が賃金を15%減少させることを示したのである。つまり、Angrist (1990) は、適切な操作変数を用いて自己選択バイアスを回避し、政策の因果効果を鮮やかに示したのである。

ラボ実験とフィールド実験 —— RCT が重要になる

属性を同じにする処置群と対照群をどうやってつくるのか、これを実験室（ラボ）で行うことは心理学では早くから行われてきた。ラボ実験（laboratory experiment）は社会科学にも導入され、たとえば、米国では主に大学生を被験者にして統制された環境で実験データを収集し、経済理論を検証する実験経済学が行われてきた。

その後、ラボを飛び出して、政策の因果効果を分析する「社会実験」とも言うべきフィールド実験（field experiment）が行われるようになる。ラボで検証された仮説や紙とペンで構築された経済学の理論が現実世界（現実世界といっても、ある時代のあつた社会という、時代的・社会的文脈に依存する限定

6) この事例は、日本経済新聞「「氷河期」100万人 就職支援」（2019年8月15日）を参考にして筆者が考えた架空の事例である。同紙によれば、政府の計画している就職支援は、民間委託型で一部成功報酬となる見込みである。「所得の引き上げで消費を底上げし、社会保障の担い手になってもらう」（同紙）ことも、この就職支援の狙いの一つである。

されたフィールドであるが)でも妥当するのかを検証するためである。とりわけ、近年の開発経済学では、フィールド実験が盛んに行われてきた⁷⁾。

ここで重要となるのが、RCT (Randomized Controlled Trial; ランダム化比較試験) という手法である。すなわち、処置群と対照群をつくる時に、RCT を用いてランダムにグループ分けすると、処置群と対照群の様々な属性を同じにすることができる。その上で、処置群に政策的介入 (treatment) を行い、処置群の結果と対照群の結果 (counterfactual potential outcome; 反実仮想的事実, 反事実の潜在的結果) の平均の差を、政策の因果効果として観測する。これは平均処置効果 (Average Treatment Effect; ATE) と呼ばれ、米国では既に一般的な政策評価の手法である。

依田 (2016: 175-176) は、かつて米国では電力需要ピーク時の節電を促すために変動型電気料金 (「ダイナミック・プライシング」と呼ばれ、電力料金が需要の少ない夜間は低くなり、需要の増大する夏の昼間は高くなる) のフィールド実験が100を超えるほど行われたが、それらのほとんどがRCT を用いていなかったことを指摘している。RCT を用いていなかったために、実験参加世帯は節電に興味のある世帯が中心となり、政策の効果が過大に推計されてしまう、自己選択バイアスが生じてしまった。このため実験結果は信頼できず、フィールド実験に投入した多額の予算が無駄になってしまったという。現在では、こうしてフィールド実験ではRCT を用いることがスタンダードになっている。

自然実験と疑似実験 — IV, RDD, DD, PSM

Angrist (1990) は、目的変数とは無関係なイベントをIV (操作変数) として利用することによって、政策の因果効果を鮮やかに示したが、これは自然災害や法改正による制度改正をIV とすることで、政策の因果効果を検証する道を拓いた。すなわち、自然災害や法改正による制度改正は目的変数と

は無関係な外生的ショック (exogenous shock) なので、適切なIV になりやすい。こうした分析手法は自然実験 (natural experiment) と呼ばれる。

社会科学では、ラボ実験やフィールド実験が行われてきた一方で、実証研究で扱われるデータは非実験的な観測されたデータ (observed data) であることの方が多い。政府統計の大半や上場企業の株価や決算情報などは実験から得たデータではないからだ。従来の計量経済学のテキストブックでは、主に観測されたデータを用いた分析手法が中心であったのはこのためである。

観測されたデータを用いる場合、処置群と対照群をデータからどうやって作るのか、そのための実証分析手法が開発・提案されてきた。たとえば、前述した操作変数法 (IV; Instrumental Variable Approach) の他に、回帰不連続デザイン (RDD; Regression Discontinuity Design), 差の差分分析 (DD; Difference in Differences), 傾向スコアマッチング (PSM; Propensity Score Matching) などである。

こうした手法は疑似実験 (quasi-experiment) と呼ばれ、RCT と並んで、因果推論 (causal inference) と呼ばれる学問分野を形成している。ビッグデータとAI (Artificial Intelligence) の活用が進む今日、因果推論は社会科学の実証研究の分析手法として注目を集めている。

4. LATE の隆盛

実証研究の地位の飛躍的向上

Hamermesh (2013) は、1963年以降に経済学の三大トップジャーナル (*American Economic Review*, *Journal of Political Economy*, *Quarterly Journal of Economics*) に掲載された論文を、その方法論によって5つに分類して、ほぼ10年おきに集計した構成比を示している。方法論の5分類とは、①理論、②シミュレーションを伴う理論、③実証

7) Duflo et al. (2008) は、開発経済学の研究成果に基づく実証分析手法の解説である。この日本語訳が監訳者による丁寧な解説付きで2019年に出版されている。また、開発経済学の実証分析については、さしあたり樋口 (2016) や伊藤 (2016) を参照されたい。

(自分で取得した訳ではないデータ), ④自分で取得したデータによる実証, ⑤実験の5つである。

Hamermesh (2013: 168, Table 4) から, その5分類の構成比を1983年→1993年→2003年→2011年の順に抜粋すると, 次のようになる。

①理論: 57.6%→32.4%→28.9%→19.1%

②シミュレーションを伴う理論:

4.0%→7.3%→11.1%→8.8%

③実証: 35.2%→47.8%→38.5%→29.9%

④自分で取得したデータによる実証:

2.4%→8.8%→17.8%→34.0%

⑤実験: 0.8%→3.7%→3.7%→8.2%

1983年以降の30年弱で, 理論研究の構成比は6割弱から2割弱に減少している。自分で取得した訳ではないデータによる実証研究は1993年までは増大しているが, 直近の2011年には3割程度までその構成比が低下している。他方, 自分で取得したデータによる実証研究は近年急速に増加し, 構成比は3割超となり, 1983年には1%未満であった実験研究の構成比は2011年には1割弱まで増加している。

このような状況は1970年代初めまでの米国の経済学会では考えられないことであった。すなわち, 1973年に米国の経済学者のアクセル・レイヨンフーヴッド (Leijonhufvud, A.) は「エコノ族の生態」という風刺論文で, 1960年代の米国の経済学者の世界を活写した⁸⁾。それによれば, 数理経済学, ミクロ経済学, マクロ経済学, 経済発展論, 実証的研究の5つの階級があり, 階級間の序列はおよそこの順であるという。少なくとも半世紀前の米国の経済学会では, 実証研究の地位が低かったことがわかる。実証研究の地位が飛躍的に向上した今日とは隔世の感がある。

LATEの旗手ともいえる経済学者である, シカゴ大学のスティーヴン・ダブナーとジャーナリス

トのスティーヴン・ダブナーの著書『ヤバイ経済学』⁹⁾の出版は2005年である。この本はベストセラーになったのだが, これは米国のジャーナリズムにおいてもLATEが市民権を得たことを示す社会現象といえるだろう。

LATEを意識した日本語テキストブックの刊行

森田 (2014) の出版された翌年以降, 我が国ではLATEで多用されている実証分析手法を扱う計量経済学のテキストブックが相次いで出版されている¹⁰⁾。

とりわけ, 田中 (2015) は学部生向けに書かれた計量経済学の入門書であるが, 操作変数法, 差の差の推定法, マッチング法, 回帰不連続デザインなど, LATEで多用されている実証分析手法が明快に解説されている。こうした手法は, これまでの伝統的な計量経済学のテキストブックには登場しなかったトピックスである。

計量経済学のみならず, LATEのトップランナーの経済学者による, ミクロ経済学やマクロ経済学のテキストブックの日本語訳も相次いで出版されている。

オースタン・ゲールズビー, スティーヴン・レヴィットおよびチャド・サイヴァーソンの3人のシカゴ大学教授によるミクロ経済学の新しい教科書も, 2017-18年に日本語訳『レヴィット ミクロ経済学』が出版されている¹¹⁾。

ダロン・アセモグル, デヴィッド・レイブソンおよびジョン・リストの3人のスター経済学者によるマクロ経済学の新しい教科書も2019年に日本語訳が出版されている¹²⁾。

さらに, 2017年にはシカゴ大学公共政策大学院の伊藤公一朗氏の書かれた一般向けの新書『データ分析の力 因果関係に迫る思考法』が出版され, サン

8) 佐和 (1982) 49-52頁。レイヨンフーヴッド「エコノ族の生態」は, *Western Economic Journal* 誌に掲載された寓話である。武藤博道氏による邦訳は筑摩書房の『展望』1974年4月号 (41-51頁) に掲載されている。

9) Levitt and Dubner (2005)。邦訳は2006年に, 増補改訂版の邦訳は2007年に出版されている。

10) 筆者の知る限りであるが, 2015年に刊行された計量経済学のテキストブックの中から例示すると, 末石 (2015), 山本 (2015), 田中 (2015), 鹿野 (2015) などが挙げられる。

11) 原書は Goolsbee et al. (2013) である。

12) 原書は Acemoglu et al. (2014) である。

トリー学芸賞や日経・図書文化賞受賞に輝いた。また、同じくベストセラーとなった、慶應義塾大学総合政策学部の中室牧子氏の『「学力」の経済学』が出版されたのは2015年である。同書はこれまで支配的であった教育評論家の言説を退け、科学的なエビデンスに基づく教育の効果を解き明かした¹³⁾。これは我が国のジャーナリズムにおいてもLATEが市民権を得始めたことを示す社会現象といえるだろう。

5. LATEの陥穽

実証分析しやすい研究テーマが選ばれる？

第3節で、Angrist (1990) が適切な操作変数を用いて政策の因果効果を鮮やかに示したことを見たが、操作変数法を用いた実証分析の肝は、いかに適切な操作変数を見つけ出してくるか、ということである。言い換えると、自然災害や法改正による制度改正のような目的変数とは無関係な外生的ショックを見つけ出してくることが、LATEにおいては重要である。

つまり、LATEの実証研究では、社会的に重要な問題からではなく、実証分析しやすい問題から研究テーマが選ばれる嫌いがある。これはLATEの陥穽といえるだろう。社会科学の社会的使命に鑑みると、これは倒錯的状况であるからだ。

もっとも研究業績を上げやすい分野に研究者が集中するのはLATEに限ったことではない。そもそも何が社会的に重要であるかは、決して価値中立的な問題ではない。

しかし、社会科学あるいは社会科学者の社会的使命については、それが社会的・時代的な文脈に依存するとはいえ（あるいは、それゆえに）、ある程度の「公約数」があるといっていよう。

難しい理論を修得しなくても、統計手法をマスターすれば論文が書ける？

LATEの優れた研究は、リサーチデザインがシンプルで直観的で分かりやすいものが多い。一流の経済学者がトップジャーナルに掲載した論文でも、必ずしも高度な経済理論が使われているわけではない。極論すれば、難しい理論を修得しなくても、統計手法をマスターすれば学術論文が書けるということになってしまう。

これは、後述するDSGEモデルなどに代表される構造推定の理論モデルが数学的に高度になりすぎて、経済学者集団の一部にしか理解されなくなっていることのバックラッシュ（反動、揺り戻し）かもしれない。そうであるならば、この点をLATEの陥穽と見るのかどうかは、多分に価値判断を伴う問題であるだろう。

LATEと構造推定は互いに補完的

こうしたLATEの陥穽に陥らないようにするためには、LATEと構造推定は互いに補完的な関係にあることに留意する必要がある。

森田 (2014) は次のように述べている。

「実は、構造推定には、それ以前に重大な短所がある。それは、自然実験に比べると、内的妥当性に劣る、という点だ。内的妥当性とは、私たちの目の前にあるデータ（あるいはその一部）からわかる範囲内については、より適切な（＝間違いの少ない、あるいは、より真実に近い）推定結果を導けることを言う。〈中略〉／これに対し、構造推定は、世界の真実の姿を非常に正確に記述したものであれば、それに基づいた構造推定は、世界の真実の姿を非常に正確に（自然実験よりも！）推定できる。〈中略〉しかし、私たちは、世界の真実の構造を決して直接知り得ないし、そのモデルが正しいかどうかも知り得ない。」¹⁴⁾

内的妥当性 (internal validity) とは、ある研究対象において説明変数と目的変数 (被説明変数) との間に因果関係があることの確からしさを示す概念である。すなわち、当該研究対象に同様の処置を

13) 同じ著者による中室・津川 (2017) もベストセラーになっているが、LATEで多用される手法 (ランダム化比較試験、自然実験、差の差分分析、操作変数法、回帰不連続デザイン、マッチング法など) が一般の読者向けにやさしく解説されている。

14) 森田 (2014) 242-243頁。

行った場合に、同様の結果が再現される程度を指す。適切なりサーチデザインのもとでシンプルな理論仮説を立てて、ある時代のある社会の特定のデータセットから処置群と対照群をつくり、その理論仮説を検証するLATEは内的妥当性が極めて高いといえるだろう。これに対して、そうして得た因果関係は、他の社会において、どの程度妥当するのかを示すのが外的妥当性（external validity）である。

一般にLATEは内的妥当性が高く、構造推定は外的妥当性が高いと言われる。これは、LATEはある時代のある社会の特定のデータセットから取り出した因果関係であるのに対して、構造推定は特定の理論モデルに依拠しているため、ある時代のある社会の特定のデータセットに依拠しているわけではないからである。

言い換えると、LATEの示す因果関係は局所的（Local）であり、一般にその外的妥当性は構造推定より低くなる。これに対して、構造推定の因果関係は大域的（Global）であり、一般にその内的妥当性はLATEより低くなる。ただし、LATEによる複数の研究を積み重ねていくと、LATEの外的妥当性は高まっていく。

ところで、近年の構造推定の代表格である動学的確率的一般均衡（DSGE; Dynamic Stochastic General Equilibrium）モデルを見ると、筆者は先に引用した森田（2014: 242-243）の指摘を痛感してしまう¹⁵⁾。

DSGEモデルに組み込まれる関数型やパラメータの符号条件等の設定如何で、さまざまな理論的帰結

や予測が導けてしまう。互いに正反対の理論的帰結や予測が導けてしまう複数のモデルが提示されても、どのモデルが現実世界をよりよく描写しているのか、どのモデルが現状分析に適しているのか、どのモデルを用いた政策シミュレーションが信頼に値するのか、客観的に判断することは困難であろう。

極論すれば、一流の経済理論家ならば、どのようなモデルでもつくれるのかもしれない。そうであるならば、現状分析や政策シミュレーションにどのモデルが適しているのかの判断は、分析者の主観に依存するだろう。

LATEと構造推定は互いに補完的ではあるが、ビッグデータとAIの時代には、LATEはますます多用されていくだろう。すなわち、ビッグデータからAIが何らかの法則性（因果関係ではなく、相関関係）を発見し、人間がその法則性を説明するシンプルな理論仮説（因果モデル）を構築して、LATEで多用されている実証分析手法で理論仮説を検証する研究が、実務でもアカデミズムでも盛んに行われるようになるだろう¹⁶⁾。

6. エビデンスをどう読むか

エビデンスの階層

政府債務の膨張によって予算制約が強まる中で、EBPM（エビデンスに基づく政策形成）は世界的な潮流になっていると言ってよい。一口にエビデンスといっても、その信頼性によって階層があることが指摘されている。

15) DSGEモデルは、オールド・ケインジアン同時方程式体系のマクロ計量モデルに対する「ルーカス批判」(Lucas 1976)に対応するために登場したモデルである。Mankiw (2010)は「ルーカス批判」について次のように説明している。「政策立案者は政策変更の効果を推計する際に、人々の期待が政策変更にどのように反応するのか知る必要がある。ルーカスは政策評価の伝統的手法——たとえば、標準的なマクロ計量モデルに依拠した手法——は政策が期待に与える影響を適切に評価していないと論じている。伝統的な政策評価に対するこの批判は、ルーカス批判として知られている」(Mankiw 2010: 451)。経済主体の期待を明示的に取り入れたDSGEモデルは、「ルーカス批判」を回避するために登場したモデルといえる。Lucas (1976)は純粋に理論的な議論なので、政策変更によって引き起こされる人々の期待の変化がどの程度のものなのか、その因果効果が究明されている訳ではない。そうした因果効果の究明は、実証研究に委ねられている。

16) ちなみに、RCTはインターネットビジネスでは既に盛んに行われており、A/Bテスト(A/B testing)と呼ばれている。A/Bテストでは、人間がまずシンプルなりサーチデザインを考え、検証するための研究仮説を立てて、その後データを集める形になる。当たり前のことだが、ビッグデータからAIがサーチデザインを決めてくれる訳ではない。伊藤(2017: 83-100)は、2008年の米国大統領選挙の際にバラク・オバマ陣営が選挙資金をインターネットで募るときに行ったA/Bテストをウェブサイト画面の写真を示して紹介している。

エビデンスの階層については、いくつかの基準が提案されている。ここでは家子ほか（2016）に基づいて、エビデンスの階層の一例を紹介したい¹⁷⁾。すなわち、エビデンスにはその質に応じて、1a から 4 までの 6 つのレベルがあり、その内容と手法の例をエビデンスの質の高い順に並べると次のようになる¹⁸⁾。

1a：RCT のメタアナリシス，系統的レビュー

【手法の例】複数の RCT のレビュー

1b：少なくとも 1 つの RTC

【手法の例】RCT（フィールド実験）

2a：少なくとも 1 つの疑似実験

【手法の例】自然実験，差の差分析，回帰不連続デザイン，操作変数法

2b：少なくとも 1 つの非実験的研究

【手法の例】回帰分析，コホート研究

3：比較試験，相関研究，記述的研究

【手法の例】前後比較，ベンチマーキング

4：専門家や実務家の見解

【手法の例】検討委員会による討議，パブリックコメント

もっとも質の高いとされるエビデンスは、RCT を用いた複数の研究成果を系統的にレビューした研究とされる。記述的研究はレベル 3 とされ、専門家の見解はもっとも質の低いレベル 4 とされている。

エビデンスの階層は、個人的経験に基づく専門家の見解がいかにかに当てにならないものか、それがフィールド実験や自然実験で検証されたエビデンスより信頼できないかを示している。

「総説」が重要になる

ところで、佐和（1982）は、我が国において 1960 年代に成し遂げられた近代経済学の「近代化」について、次のように描写している。

「米国のすぐれた「教科書」が陸続と翻訳出版され、幸か不幸か学生は、経済学を学ぶのに「古

典」をひもとく必要を感じなくなった。専門の経済学者は、米国の雑誌に論文を掲載することを、みずからの身上と心得るようになった。〈中略〉アメリカ経済学会の近代的運営方式は、ことごとく日本の学会にも実現された。学会誌にはレフェリー制が導入され、老大家が書く長大なサーヴェイ（展望）論文は、独創性を欠くという、しごくもつともな理由で学会誌への掲載を拒絶されるようになった。」¹⁹⁾

我が国では、1960 年代の近代経済学のみならず、その後の社会科学系の学会では、おおよそ上記のような経緯で、学会誌に掲載される論文は原著論文が中心となり、サーヴェイ論文は締め出されてきた嫌いがある。

しかし、学術研究の中心が理論研究ではなく因果推論の実証研究になってくると事情は変わってくる。すなわち、様々な因果関係を示すエビデンスを提示する原著論文が陸続と公刊されると、同じテーマでも因果関係があるという論文と、因果関係がないという論文が登場してくる。仮に分析手法に問題がなくても、分析したデータセットが異なるため、こうした相対立する分析結果が出ることは決して不思議なことではない。

分析したデータセットを生み出した母集団は何か、検証された仮説の外的妥当性はどこまで通用するのか、より高次の観点から複数の研究結果を検討・分析して、全体としてどのような因果関係があるのかを明らかにする必要がある。

自己の主張に都合のいい論文ばかりを集めて自説を主張する「チェリー・ピッキング」を慎み、科学的に確立された手続きで先行研究を纏めるメタアナリシスが重要になる²⁰⁾。

かつて学会誌から締め出された「総説」が、エビデンスとしての質がもっとも高いものとして、再び注目される時代になったといえるだろう。

17) 本稿で紹介するエビデンスの階層は、家子ほか（2016: 4）に登場する「図表 2 各種手法のエビデンスレベルと具体例」である。これは三菱 UFJ リサーチ & コンサルティングが惣脇（2010）等を参考に作成したものである。

18) 家子ほか（2016）4 頁。なお、「手法の例」の術語の日本語表記は一部修正した。

19) 佐和（1982）130 頁。

20) 中室・津川（2017）74-76 頁。

7. 結びに代えて

本稿で見てきたように、「信頼性革命」は確かに大きな実証分析手法の刷新ではあるが、それはトーマス・クーンの意味での「科学革命」ではないといえるだろう²¹⁾。社会科学者集団がこれまでのパラダイムを捨てて、新しいパラダイムに乗り換えた訳ではないからだ。

実証研究の分析手法として、LATEと構造推定は互いに補完的な関係にある、という捉え方が一般的になりつつある²²⁾。

実証分析手法の進歩は日進月歩であり、社会科学や人文科学の分野にも、いわゆる理系の分析手法や理論が導入されて、既成の学問が塗り替えられている。

橘(2012)は「社会科学・人文科学を進化論的に基礎づけるパラダイムシフト」²³⁾がおきていることを指摘している。ここでいう知のパラダイムシフトとは、橘氏の言葉を借りれば次のようなものである。

「1980年代から、知の世界では大きな地殻変動が起きている。それは端的にいえば、ダーウィンの進化論によって人間の心理や行動、社会の仕組みを説明しようとする進化心理学と、脳の構造を解明する脳科学が融合して、経済学や社会学、政治学、さらには哲学や倫理学に至るまであらゆる社会科学・人文科学を進化論的に基礎づけるパラダイムシフトのことだ。」²⁴⁾

EBPM(エビデンスに基づく政策形成)や知のパラダイムシフトの潮流の中で、地域政策学も今後大きく変貌していこう。

筆者は細分化と専門化が高度に進んだ現代科学において、学際融合や専門知の統合が新たな知を生み出すものと思う。それはクーンの意味での「科学革命」を引き起こすかもしれない。

(参考文献)

【英語文献】

- Acemoglu, D., Laibson, D. and List, J. (2014) *Macroeconomics*, Pearson. [岩本康志監訳(2018)『アセモグル/レイブソン/リスト マクロ経済学』岩本千晴訳, 東洋経済新報社。]
- Angrist, J. D. (1990) Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records. *American Economic Review*, 80(3): 313-336.
- Angrist, J. D., and Pischke, J. (2009) *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press. [大森義明・小原美紀・田中隆一・野口晴子訳(2013)『「ほとんど無害」な計量経済学——応用経済学のための実証分析ガイド』NTT出版]
- Angrist, J. D., and Pischke, J. (2010) The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. *Journal of Economic Perspectives*, 24(2): 3-30.
- Duflo, E., Glennerster, R. and Kremer, M. (2008) Using Randomization in Development Economics Research: A Toolkit. Schultz, T. P. and Strauss, J. (eds.) (2008) *Handbook of Development Economics*, Vol.4, Part 13, Chapter 61, pp.3895-3962, North Holland. [小林庸平監訳・解説(2019)『政策評価のための因果関係の見つけ方——ランダム化比較試験入門』石川貴之・井上領介・名取淳訳, 日本評論社]
- Goolsbee, A., Levitt, S. and Syverson, C. (2013) *Microeconomics*, Worth Publishers. [安田洋祐監訳(2017; 2018)『レヴィット ミクロ経済学 基礎編・発展編』高遠裕子訳, 東洋経済新報社]
- Hamermesh, D. S. (2013) Six Decades of Top Economics Publishing: Who and How? *Journal of Economic Literature*, 51(1): 162-172.

21) トーマス・クーン (Kuhn 1962) の「科学革命」論については、佐和 (1982) 154-155頁が極めて明快に解説している。

22) たとえば、森田 (2014) や中嶋 (2016) や Heckman (2010) を参照されたい。

23) 橘 (2012) 380頁。

24) 橘 (2012) 380頁。橘氏の一連の著作において、現代の進化論をベースに展開される各学問分野の記述については、通説になっていない議論も含まれる。それらに対して各分野の専門家から異論があることも、筆者は一部承知している(橘氏も承知していることだろう)。科学の最先端において、新たなエビデンスによって通説が塗り替えられていくことは、クーンのいう通常科学の営みにおいても珍しいことではない。

- Heckman, J. J. (2010) Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy. *Journal of Economic Literature*, 48(2): 356-398.
- Kuhn, T. S. (1962) *The Structure of Scientific Revolutions*, University of Chicago Press. [中山茂訳 (1971) 『科学革命の構造』 みすず書房]
- Levitt, S. D. and Dubner, S. J. (2005) *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*, William Morrow. [望月衛訳 (2007) 『ヤバい経済学——悪ガキ教授が世の裏側を探検する』 増補改訂版, 東洋経済新報社]
- Lucas, R. E. (1976) Econometric Policy Evaluation: A Critique. In Brunner, K. and Meltzer, A. (eds.) *The Phillips Curve and Labor Markets. Carnegie-Rochester Conference Series on Public Policy*, 1(1): 19-46, American Elsevier.
- Mankiw, N. G. (2010) *Macroeconomics*, 7th ed., Worth Publishers. [足立英之・地主敏樹・中谷武・柳川隆訳 (2011; 2012) 『マンキュー マクロ経済学 入門篇・応用編』 第3版, 東洋経済新報社]
- 【日本語文献】**
- 樋口裕城 (2016) 「開発経済学における計量的アプローチと実験的アプローチ」 経済セミナー増刊号「進化する経済学の実証分析」2016年9月20日, 63-69頁, 日本評論社。
- 依田高典 (2016) 『「ココロ」の経済学——行動経済学から読み解く人間のふしぎ』 筑摩書房。
- 家子直幸・小林庸平・松岡夏子・西尾真治 (2016) 「エビデンスで変わる政策形成——イギリスにおける「エビデンスに基づく政策」の動向, ランダム化比較試験による実証, 及び日本への示唆」『政策研究レポート』三菱UFJリサーチ&コンサルティング, 2016年2月12日。
http://www.murc.jp/thinktank/rc/politics/politics_detail/seiken_160212.pdf (2017年4月30日閲覧)
- 伊藤公一朗 (2017) 『データ分析の力 因果関係に迫る思考法』 光文社。
- 伊藤成朗 (2016) 「開発経済学」 経済セミナー増刊号 (前掲), 102-113頁。
- 鹿野繁樹 (2015) 『新しい計量経済学 データで因果関係に迫る』 日本評論社。
- 北村行伸 (2016) 「応用ミクロ計量経済学の手法と論点」 経済セミナー増刊号 (前掲), 28-36頁。
- 森田果 (2014) 『実証分析入門——データから「因果関係」を読み解く作法』 日本評論社。
- 中室牧子 (2015) 『「学力」の経済学』 デイスクヴァー・トゥエンティワン。
- 中室牧子・津川友介 (2017) 『「原因と結果」の経済学——データから真実を見抜く思考法』 ダイヤモンド社。
- 中嶋亮 (2016) 「「誘導型推定」 v.s. 「構造推定」」 経済セミナー増刊号 (前掲), 52-61頁。
- 澤田康幸 (2016) 「経済学における実証分析の進化」 経済セミナー増刊号 (前掲), 13-19頁。
- 惣脇宏 (2010) 「英国におけるエビデンスに基づく教育政策の展開」『国立教育政策研究所』 第139集, 153-168頁。
https://www.nier.go.jp/kankou_kiyou/kiyou139-016.pdf (2017年4月30日閲覧)
- 佐和隆光 (1982) 『経済学とは何だろうか』 岩波書店。
- 末石直也 (2015) 『計量経済学——ミクロデータ分析へのいざない』 日本評論社。
- 橘玲 (2012) 『(日本人)』 幻冬舎。
- 田中隆一 (2015) 『計量経済学の第一歩——実証分析のススメ』 有斐閣。
- 渡部敏明 (2016) 「ルーカス批判とマクロ計量分析」 経済セミナー増刊号 (前掲), 37-41頁。
- 山本勲 (2015) 『実証分析のための計量経済学』 中央経済社。