



シリーズ
日本経済を考える

68

「エビデンスに基づく政策形成」とは何か

前財務総合政策研究所研究官
山名 一史*1

1. はじめに

巨額な公的債務を抱える我が国において、効率的な資源配分は重要な課題であり、「政策にどれだけの予算がかけられたか」ではなく、「政策目的がどれだけ効率的に達成されたか」という観点から、政策は評価されなければならない。こうしたニーズに応えるため、近年注目を集めているのが「エビデンス（客観的根拠）に基づく政策形成（Evidence-based policymaking：EBP（M）」である（例えば、黒田・有本（2010）、家子他（2016）、森川（2016）、藤田編（2016）、EBPMのニーズに対応する経済統計の諸課題に関する研究会（2016））。つまり、効果が合理的に予測される政策のみを行うべきである、という考え方である。治療を行う際は、個人の経験や勘に基づく治療法ではなく、効果があると科学的に認められた治療法を選択するべきである*2、と考えれば分かりやすい。本稿では、エビデンスをどのように得るか、という手法的な観点からEBPに関する議論を整理する。本稿の構成は以下の通りである。まず第2節で国際機関や諸外国がEBPにどのように取り組んでいるかを概観し、第3節ではエビデンスの定義と代表的な分析手法を紹介し、そ

れぞれの長所と短所を考察する。最後に第4節を本稿のまとめとする。

2. 国際機関や諸外国のEBPへの取り組み

日本では比較的新しい概念のEBPであるが、国際機関や先進諸国において、政策形成におけるエビデンス重視の流れは既に一般的である。例えば、OECDの教育研究革新センター（CERI）では2003年からエビデンスに基づく教育政策の研究に取り組んでおり、2007年には報告書“Evidence in Education：Linking Research and Policy”が公表されている*3。また、2015年に開催された第3回IMF統計フォーラム“OFFICIAL STATISTICS to Support Evidence Based Economic Policymaking”*4では、エビデンスに基づく政策形成と公的統計マイクロデータとの関連について幅広い議論が行われている。さらに世界銀行では、開発インパクト評価（Development Impact Evaluation：DIME）により、エビデンスに基づく公共政策の企画立案・実施を推進している*5。米国においても、2001年にブッシュ政権下で可決された「落ちこぼれ防

*1) 本稿の執筆にあたって、財務省財務総合政策研究所の大関由美子財政経済計量分析室長、山崎丈史主任研究官、研究員の皆様より有益な助言や示唆をいただいた。なお、本稿の内容や意見は全て筆者の個人的な見解であり、財務省および財務総合政策研究所の見解を示すものではない。ただし、本稿における誤りのすべては、筆者の責に帰すべきものである。

*2) 保健医療分野では、このような考え方を、既存の経験と勘に基づく医療と区別するため、エビデンスに基づく医療（Evidence-based medicine、EBM）と呼んでいる。

*3) <https://www.oecd.org/edu/ceri/evidenceineducationlinkingresearchandpolicy.htm>

*4) <https://www.imf.org/external/np/seminars/eng/2015/statsforum/>

止法」(No Child Left Behind Act : NCLB) では、科学的根拠のある研究 (Scientifically-based research) という文言が条文内に100回以上登場するなど、科学的根拠に基づく施策を重視する姿勢が顕著であり (田辺 (2006))、こうした姿勢はオバマ政権下で可決された「エビデンスに基づく政策形成のための評議会法」(Evidence-Based Policymaking Commission Act of 2016)*6にも反映されている (伊藤 (2016))。最後に、英国ではブレア政権下の1999年以降、エビデンスを活用した政策の推進に継続的に取り組んできた歴史があり、What Works Centre (WWC) と呼ばれる官民協働組織がその中核を担っている (家子他 (2016)・小林 (2016))。翻って、我が国においては医療研究分野や科学技術政策など一部の分野でEBPへの取り組みが見られる (黒田・有本 (2010)) もの、大規模な実証研究の蓄積やEBPを担保するための仕組みづくりは進んでいない。

3. EBPに必要な分析手法

3-1. エビデンスとは何か

EBPを理解するためには、まずエビデンスの定義を確認する必要がある。「エビデンス=客観的根拠」という言葉を文字通り受け取ると、統計調査によって得られた数字は客観的なもので、これすなわちエビデンスだと言えそうだ。しかし、EBPの文脈でエビデンスという場合、そうした単純な数字の羅列では不十分、政策と政策効果との間に見られる相関関係でも不十分で、因果関係が証明されて初めてエビデンスと呼ばれるようになる。なぜなら、相関関係は因果関係の必要条件であって十分条件ではないからである。因果関係ではないにもかかわらず相関関係が観察される例として、逆の因果関係、交絡変数 (交絡因子、潜伏変数)*7の無視、偶然の産物などが挙げられ、これらは擬似相関と総称される。

エビデンス=客観的に示された因果関係、という図式を理解すると、なぜ相関関係では不十分なのかという素朴な疑問が生じる。その理由を端的に説明すれば、擬似相関に基づく政策は好ましくない効果を生むからであり、これは具体例で確認することが出来る。

第一に、犯罪数と警官数との間に正の相関関係が観察される例を考えよう。実際は犯罪の増加に対応するために正相関が観察されているのだが、ここで政策担当者が「警官数が増えたから犯罪数も増えたのだ」という因果関係を誤認すると、犯罪数を減少させるために「警官数を減らす」という政策が推奨されることになる。結果的に警官数の減少は抑止力の低下を通じて犯罪数を増加させることになり、期待している効果とは真逆の政策効果が生まれる。

第二に、アイスクリームの売上高と殺人件数との間に正の相関関係が観察される例を考えよう。この正相関は、売上高と殺人件数とに共通する「気温」という要素 (=交絡変数) を無視しているために観察されているものだが、ここで政策担当者が「アイスクリームに含まれる危険物質が殺人を誘発している」という因果関係を誤認すると、殺人件数を減少させるために「アイスクリームの販売を規制する」という政策が推奨されることになる。言うまでもなくアイスクリームの販売規制は殺人件数に何の影響も与えない*8ため、結果的に無駄な政策に予算をつぎ込むことになる。

第三に、地球温暖化と海賊の数との間に負の相関関係が観察される例を考えよう。地球温暖化と海賊の数との負の相関は偶然観察されているに過ぎないが、ここで政策担当者が「海賊数が減ったから地球が温暖化した」という因果関係を誤認すると、温暖化を食い止めるために、「海賊を増やす」政策、つまり海賊行為がし易いように武装警備員の乗船許可を認めない、などの措置が推奨さ

*5) <http://www.worldbank.org/en/research/dime>

*6) <https://www.congress.gov/bill/114th-congress/house-bill/1831>

*7) 非説明変数と説明変数の両方に相関する変数のこと。後述の具体例で説明する。

*8) 規制に怒りを覚えたアイスクリーム屋や消費者が、軽犯罪に走るといった負の効果はありそうである。

れることになる。結果は明白で、海賊の被害者が激増する一方で温暖化は進行し続けることになり、期待している政策効果がない上に意図しない被害が生まれることになる。

3-2. 政策評価法

それでは、どのような手法を用いれば相関関係ではなく、因果関係を確かめられるだろうか。EBPの文脈で最初に紹介されるのが、政策評価法(Program evaluation (Imbens and Wooldridge (2009))), より具体的にはランダム化比較試験(Randomized control trial : RCT) と準ランダム化(Quasi-randomization) という2種類の因果推定法である。政策評価法は既存の誘導型推定(Reduced-form estimation)、すなわち経済理論を単純な計量モデル(例えば回帰分析や後述のVAR(ベクトル自己回帰モデル))で近似し、得られた変数間の相関を経済理論に基づいて解釈するアプローチでは、疑似相関に起因した政策形成の誤りが生じる、という反省から発展した分析方法である*9。

3-2-1. RCTを用いた政策評価とその問題点

RCTは結果に影響を与える「他の条件が一定」になるような状況を「人為的」に作ることで、実験計画法に基づいて因果関係を検証する方法であり、メタ分析*10に次いで厳密な因果検証方法である。ウェブマーケティングの分野ではA/Bテストという名で用いられる。RCTでは、まず実験を行うサンプル(マウスや人間など)をランダム(無作為)に2つの群に割り当て、一方の群には薬品の投与を、もう一方の群には何も投与しない*11。ここで投与を受ける前者を処置群(実験群)、投与を受けない後者を統制群(参照群)と

呼ぶ。そして、投与した後に2つの群の反応を比較することで、薬品の効果を明らかにする。この時、投薬以外の条件がすべて一定なので、生まれた反応の差は投薬によるものだと結論付けられる。

RCTを行う際にもっとも重要なのは、サンプルを「ランダム」に2つの群に割り当てる点である。もしランダムに割り当てずに分析を行うと、2つの群の反応差が薬効によるものか、それとも群の属性の偏りによるものか識別できない、いわゆるサンプル・セレクション・バイアスが発生し、比較結果の解釈が困難になる。このバイアスの例としてよく用いられるのが、申請方式の職業訓練プログラムである。失業者への職業訓練プログラムがどの程度再就職率を改善させるか、その効果を知りたいとしよう。もし、プログラムを申請方式にして実験を行うと、相対的に意欲の高い失業者が申請を行って処置群になるため、処置群と統制群との再就職率の差がプログラムによるものか、それとも意欲の差によるものか識別が困難になる。同様のバイアスは、新しい教育プログラムを申請方式で導入すると、教育意欲の相対的に高い教師のクラスが処置群になる場合に発生するだろうし、研究費を公募方式にすると、意欲が相対的に高い研究者が処置群になる場合にも発生するだろう。もちろん意欲が評価されること自体は悪くないが、プログラムの因果効果を検証するという観点からは、こうしたバイアスを避ける必要がある。たとえばA/Bテストでは、同時点でウェブサイトを訪れた訪問者に対し、ランダムに2種類のデザインのページを表示することで、ランダム性を確保している。

RCTは因果効果を推定できる強力な手法であるが、政策効果の予測に使うには幾つかの問題点

*9) 政策評価法でない単純な誘導型推定では厳密なエビデンスを得られないが、幅広い問題を取り扱うことが可能であるため、政策効果の計測や予測方法として伝統的に用いられてきた。

*10) メタ分析とは、多数の独立なRCTの結果を統合して行う分析のことである。

*11) ただし、実験対象が人間の医療試験の場合はプラセボ(偽薬、プラシーボともいう)効果が存在するので、統制群にプラセボを与える場合がある。プラセボを与える統制群と区別するため、何の処置も行わない統制群を無処置群と呼ぶことがある。また、統制群にプラセボを与える試験をプラセボ対照試験と呼び、統制群が無処置群である時の無治療対照試験と区別することがある。プラセボ対照試験のように、実験の効果を実験対象から分らないようにして行う試験法を単盲検法と呼ぶ。

がある。第一に、費用など物理的な制約の問題があげられる。A/Bテストのような小規模実験は、比較的容易に実装可能であり、費用対効果が高いため、ウェブマーケティングで頻繁に活用されている。他方、職業訓練プログラムや教育プログラムといった政策の因果効果を分析するには大規模な社会実験が必要であり、予算制約や費用便益の観点から必ずしも実行可能ではない。また、処置のランダムな割り当てが物理的に困難、または倫理的に許容できない状況も少なくない。

第二に、結果を解釈する際に生じるバイアスの問題があげられる。RCTはその性質上、莫大な費用と時間が投資されるため、実験結果を論文として出版したいという研究者のインセンティブが相対的に強く、政策効果に対して好意的な結果が報告される傾向が高い。このようなバイアスは観察者バイアスや出版バイアスと呼ばれ、統計的に有意な結果が得られるまでデータ収集や統計分析を続ける研究者の慣習はP-hacking (Head et al. (2015)) と呼ばれる。観察者や分析者に起因するこうしたバイアスは一般的に観察されるため、例えば、医療保険分野では、二重盲検法を用いることで観察者バイアスに対処している。ここで二重盲検法とは、単盲検法のように実験の効果を実験対象から分からないようにするだけでなく、観察者からも分からないようにして実験を行い、結果の恣意的な解釈を防止する方法である。

最後に、RCTが分析可能な政策やRCTによって得られる政策含意は限定的である点に注意が必要である。RCTの結果が意味を持つためには「処置群が受ける処置は、統制群に影響を与えない」というSUTVA (Stable-Unit-Treatment-Value-Assumption) (Rubin (1978)) と呼ばれる仮定が満たされなければならない。しかし、現実にはSUTVAが満たされる状況はあまり多くない。具体例として、職業訓練プログラムを再考しよう。仮にプログラムが失業者に対してランダムに割り当てられたとしても、そもそもの求人数には限り

がある。80人の求人募集に対して100人が応募するような状況で、失業者を無作為に2つの群に分けると、平均40人がそれぞれ再就職できる。ここで片方に訓練プログラムを行った結果、処置群の再就職数が40人から60人に増加したとしよう。確かにこの増加はプログラムの効果かもしれないが、同時にもう一方の群、すなわち統制群の再就職数は40人から20人に減少する。明らかにこの統制群は訓練プログラムから負の影響を受けており、SUTVAが満たされていない。

SUTVAの議論からは二つの含意を得ることが出来る。第一に、SUTVAが満たされていなければ処置効果が意味を持たないので、RCTで因果効果を検証することが出来ない。第二に、SUTVAが満たされた状況でRCTによって検証された因果効果は一般化することが出来ない。なぜなら、一般的な環境下ではSUTVAが保証されないからである (Imbens and Wooldridge (2009), Heckman et al. (1999))*¹²。第二の点は実験の外的妥当性 (External validity)*¹³として議論されており、伝統的な経済学の用語で表現すれば、「RCTで検証できるのは部分均衡における因果効果に限られる。」ということになるだろう。これは政策担当者にとって致命的な欠点になりうる。何故なら、予算の効率的な配分の観点から考慮すべきはマクロ的な一般均衡効果であって、ミクロ的な部分均衡効果ではないからである。

3-2-2. 準ランダム化を用いた政策評価とその問題点

RCTは「実験計画法に基づいて、今後実行する政策のエビデンス」を蓄積していく方法であり、「実験計画法に基づかずに実行された過去の政策に関する観測データ」からエビデンスを得る方法ではない。しかし、政策担当者の実務的なニーズに応えるのはむしろ后者である。このような動機から用いられるのが準ランダム化と総称される手法で、具体的には自然・疑似実験 (準実験)

* 12) 10,000人の求人募集に対して100人のRCTであれば、統制群はしわ寄せの影響を受けないのでSUTVAに違反しない。しかし、この結果を労働市場全体に一般化することはできない。

* 13) GeneralizationやExtrapolationと表現されることもある (Deaton (2010))。

アプローチである*14。準ランダム化では、あたかもランダムに政策の影響が割り当てられ、実験的な状況が作り出されたとみなすことができる過去の事例を探し出し、因果効果の検証を行う。これらの方法で使われる手法の内、本稿では差の差推定と操作変数法を紹介する*15。

いま、偶然に政策の効果を受けた（ランダムに政策効果を割り当てられた）とみなすことが出来る集団のデータが得られたとしよう。この集団は処置群と考えることができ、政策前後での集団の反応の差（ d_T ）は政策から受ける影響（処置効果（Treatment effect）と呼ばれる、 d_p ）とそれ以外の要素から受ける影響（ d_o ）に分解することが出来る（ $d_T = d_p + d_o$ ）。ここで、政策から一切の影響を受けず（ $d_p = 0$ ）、かつそれ以外の要素からは処置群と同じ影響を受ける集団、すなわち統制群を選ぶ（これをマッチングと呼ぶ）ことができたとしよう。統制群の政策前後での反応の差（ d_c ）は処置効果を受けていないので、 $d_c = d_o$ となる。最後に、「2つの集団の前後の反応の差」の差分を取れば、処置効果、つまり政策の因果効果を抽出することが出来る。これが差の差（Difference in Differences：DD又はDID）推定と呼ばれる考え方である（ $d_T - d_c = (d_p + d_o) - d_o = d_p$ ）*16。

言うまでもなく、差の差推定で登場するような都合のいい処置群と統制群の組み合わせは都合よく見つからない。しかし、交絡変数と独立で、かつ説明変数と相関する変数なら見つかることがあるかもしれない。もし、こうした変数が見つかるのであれば、交絡変数の影響をうまく取り除いて擬似相関の疑いを晴らすことができるため、因果推定を行うことが可能となる。そこで、このような変数を操作変数（Instrumental variable：IV）と呼び、IVを用いた因果推定をIV法と呼ぶ。

代表的なIV法を用いた実証研究は、従軍経験が生涯所得に与える影響を調べた Angrist (1990) である。一般に従軍するかどうかの意思決定はランダムに行われなかったため、仮に従軍経験と生涯所得との間に正の相関が見られたとしても、従軍経験が生涯所得に正の影響を与えたのか、それとも従軍するかどうかの意思決定に影響を及ぼし、かつ生涯所得にも正の影響を持つような別の要因、いわゆる交絡変数が正相関を生み出しているのかを識別できない。こうしたサンプル・セレクション・バイアスを除去するため、ベトナム戦争中の1970年から1971年には誕生日くじ（Lottery）によってランダムに徴兵が行われた期間がある、という事実に着目したのが研究のポイントである。誕生日くじによる徴兵は、説明変数である従軍経験と関連するが、交絡変数とは独立である。そこで、誕生日くじによる徴兵が従軍経験に与える影響に対応して、どの程度生涯所得が変動するかを測ることで、交絡変数の影響を取り除き、従軍経験が生涯所得に与える因果効果の推定を行うのがIV法の考え方である*17。

一見すると、準ランダム化は「過去の観測データ」から因果効果を抽出できる魔法の杖にも見えるが、現実にはRCTと同様、様々な制約が存在する。例えば、準ランダム化の文脈において観測データとは、同一の対象を継続的に観察し、記録したパネルデータを意味しており、パネルデータが得られないのであれば、一般に因果推定は不可能である。また、仮にパネルデータが得られたとしても、ランダムに政策効果が割り当てられたと見なせる集団や都合のいい代理変数（Surrogate）、またはIVが見つかる可能性はきわめて低い。準ランダム化は、あくまでも理想的な環境があつて初めて検討すべき手法であつて、割り当てのランダム性に疑問が生じるような状態で使うべきでは

* 14) 人為的に純粋なランダム化を行うRCTと区別するため、“準”ランダム化と呼ばれる。

* 15) 他にも部分識別を用いる方法や回帰不連続デザインを用いる方法など、様々な分析手法が考案されている（邦語文献では中室・津川（2017）、田中（2015）、鹿野（2015）、星野（2009）などを参照のこと）。

* 16) 実際は個々の主体によって効果の程度が異なる（異質性）ため、平均処置効果（Average treatment effect：ATE）を計算する。

* 17) ただし、誕生日くじによる徴兵は必ずしも従軍を意味しない（徴兵拒否が存在する）ので、実際にはより詳細な議論が必要となる。

ないし、逆にランダムな割り当てが認められるような都合のいい集団や代理変数を偶然見つけたからといって、そこから得られる些末な因果効果の分析を行うのは本末転倒である。ハイゼンベルクとの対話でアインシュタインがいみじくも語っているように、「原理的な観点からは、観測可能な量だけをもとにしてある理論を作ろうというのは、完全に間違っています。なぜなら実際はまさにその逆だからです。理論があつてはじめて、何を人が観測できるかということが決まります。」(ハイゼンベルク (1974), p104)。データはあくまでも理論的な予測の検証にのみ使うべきである。

準ランダム化を用いる際のその他の問題点として、RCTと同様、出版バイアスの問題があげられる。特に、準ランダム化はRCTと比較して群の選択に研究者の恣意性が入りこむ余地が多い。誤った処置群の選択はサンプル・セレクション・バイアスを生むことに十分留意するとともに、少なくとも変数選択はSpike and slab事前分布を用いた正則化(罰則付き)回帰で自動化する*18など、分析者の恣意性が入り込む余地を最小化する努力が求められる*19。また、検証された因果効果が処置群と統制群の選択にどの程度依存しているのか、その頑健性は入念に検証されなければならない。最後に、得られた因果効果は、RCTと同様、部分均衡的な因果効果に過ぎない。

3-3. 準エビデンス

これまで見てきたように、RCTは「過去の観測データ」からは何も言えず、準ランダム化がデータから因果関係を抽出できるのは偶然の産物に過ぎない。しかし、実務では「必ずしも頑健でなくてもいいので何らかの政策根拠」が必要なこと

がある。このような動機から、政策効果の計測や予測に用いられてきた統計手法の一つがVAR(ベクトル自己回帰)モデルである。VARは、内生変数の動きを内生変数のラグと外生変数であるショック項とでモデル化する。ここで、仮に構造VAR、すなわち経済理論から予想される変数間の関係を同時点制約として課したVARを用いたとしても、推定される関係は相関関係に過ぎず、EBPが要求する水準のエビデンスが得られるわけではない*20。しかし、多数の独立なVARの結果を統合してメタ分析を行うなど追加的な検証を行えば、過去の政策効果について限りなくエビデンスに近い結果を得ることが可能かもしれない。そこで、本稿ではこのような限りなく因果関係に近い相関関係を「準エビデンス」と呼ぶ。準エビデンスは一般均衡効果を考慮でき、さらにデータ制約に直面しにくい様々な過去の政策効果を分析できるが、あくまでも相関に過ぎず、厳密な因果性との間には越えられない壁が存在する。

3-4. 政策効果の予測

さて、ここまで「過去の政策効果の有無」に關する手法を紹介してきた。なぜなら、「過去に政策効果が認められた(認められなかった)」政策を行った場合、「同様の効果が予測される(予測されない)」とEBPが暗黙に仮定しているからである。確かに、こうした議論は医薬品や治療法の効果については当てはまりそうであるが、経済政策についても同様に、過去に効果があったという事実から将来の政策効果を予測するのは合理的だろうか。

予測の妥当性に関して、まず考慮すべきなのが「ルーカス批判」(Lucas critique)である*21。ルーカス批判とは、「政策を変更すると、家計や企

*18) Spike and slab事前分布とは、あるパラメータについて、そのパラメータが非ゼロである確率の事前分布(Spike)と、どのような値を取るかという事前分布(Slab)の混合分布を考えることを意味する。また、正則化回帰とは二乗誤差関数に罰則(ペナルティ)項を加えて最小化を行うことで、パラメータの推定と変数選択を同時に行う手法である(Lassoや、LassoとRidge回帰を組み合わせたElastic netなどがよく用いられる)。詳しくはFriedman(2010)などを参照のこと。

*19) 言うまでもなく、前提となる変数の選択は経済理論に基づいて検討されていなければならない。

*20) もっとも、相関は因果でないが役に立つ(場合が少なくない)。例えば、最小分散ポートフォリオの設計時、資産ごとの最適な投資比率は過去の相関の情報を用いて計算される。

業といった経済主体の意思決定が変化するため、過去のデータから推定された（誘導型）パラメータを用いた政策効果の予測は無意味である」というものである。例えば変数 X と Y の間に線型関係 $Y = \beta X$ があり、過去のデータの時系列から $\hat{\beta} = 2$ であることが推定できたとしよう。ここで、変数 X を増やす政策（ ΔX ）の効果も、 $\Delta Y = 2\Delta X$ としてはいけないというのが、ルーカス批判の意味するところである。なぜなら、 β は内生的に値が決まるパラメータであり、政策によってその値は変化し得るからである。

それでは、これまで紹介してきた手法を用いた予測はルーカス批判を免れるだろうか。結論から言えば、どの手法を用いても将来行う政策の有効性は保証されない、というのがルーカス批判の理論的な帰結である。なぜなら、単純な誘導型推定によって導かれる相関関係や準因果関係については言うまでもなく*22、RCTや準ランダム化によって認められた頑健な因果関係であっても、経済主体の意思決定が変われば、観察されてきた変数の関係は崩れるからである*23。ルーカス批判を考慮すると、RCTや準ランダム化によって得られる「過去の政策効果に関するエビデンス」と、EBPが本来必要としている「将来の政策効果を合理的に予測するためのエビデンス」との間には、大きな隔たりが存在するのである*24。

では、どのようにすればEBPが必要とする「将来の政策効果を合理的に予測するためのエビデンス」を得られるだろうか。これまでの議論を踏まえると、エビデンスはルーカス批判に抵触せず、一般均衡効果を考慮できるフレームワークで、

「実験計画法に基づかずに得られた過去の観測データ」から推定される因果関係であることが望ましい。こうした要請に応えることができる手法として、構造推定（Structural estimation）と状態空間モデルを用いた因果推定という2種類の方法があげられる。どちらもルーカス批判に抵触しないが、後者は将来の予測に用いることが出来ないため、準ランダム化の欠点を補完する手法という位置づけである。本稿ではEBPと密接に関連する構造推定について説明を行う*25。

3-4-1. 構造推定

構造推定とは、経済主体の（動学）最適化行動を明示的に考慮した一般均衡（部分均衡）モデルを構築し、構造モデル内のパラメータを推定する手法である。なぜ、構造推定によって推定されたパラメータを用いた予測は、RCTや準ランダム化から得られたエビデンスを用いた予測と違ってルーカス批判に抵触しないのだろうか。それは、前者のエビデンスは「政策変更によって可変」なパラメータに基づいているのに対し、構造モデル内のパラメータは経済主体の選好や技術水準のように根源的（Primitive）なものであり、「政策変更によって不変」だからである*26。そして、根源的なパラメータと因果関係が明示的に規定された理論モデルを用いて反実仮想シミュレーション（Counter-factual simulation）を行えば、これまでに行ったことのある政策効果の測定だけでなく、これまでに行ったことのない政策効果の測定も可能になるというのが構造推定を用いた予測の最大の長所である（Nevo and Whinston (2010)）。

*21) Lucas (1976)、邦語文献では、例えば渡部 (2016) を参照。

*22) ルーカス批判とは無関係の、疑似相関に起因する予測の誤りが存在するため。

*23) 同様の政策を行えば、同じように経済主体は反応するはずだ、と考えることも出来なくはない。しかし、他の条件が異なる環境下で、ある政策の変化に対して「経済主体が以前の政策変更時と同様にふるまう」と信じる根拠は存在しない。

*24) だからといって、「過去の政策効果に関する部分均衡的なエビデンス」が無意味なわけではない。過去の政策効果の測定は、経済学の重要なテーマの一つである。過去の政策評価は重要だが、その結果を予測に用いてはならない、と述べているに過ぎない。

*25) 後者は、機械学習（Machine learning）分野において、ベイズ構造時系列モデル（Bayesian structural time-series model）を用いた因果推定とも呼ばれる。ここでいう「構造」とは、構造推定の「構造」、すなわち「経済主体の最適化問題を明示的に考慮する」の意味ではなく、「トレンド項や季節調整項といった要素をモデル内で明示的に考慮する」の意味で用いられる。詳しくはBrodersen et al. (2015) を参照のこと。

確かに、構造推定は経済理論に基づく定式化に結果が大きく依存しており、信頼性に欠けるといふ批判も存在する (Angrist and Pishke (2010))。しかし、統計モデルは程度の差こそあれ、何らかの制約を置いて分析を行っており、「構造推定の制約は過剰であるが誘導型推定の制約は許容される」といった判断は恣意的である。また、政策評価法の仮定は少ないように見えるが、実際は構造推定の仮定のように明示されていないだけで暗黙の仮定が多く、実質的にどちらがより制約的かは自明でない。仮に上記の批判がすべて当てはまるとしても、構造推定はルーカス批判に対応した予測が出来る唯一の手法であり、特に政策の事前評価段階で積極的な活用が期待される。

元々、「エビデンスに基づく」という表現は、構造推定のように経済理論を仮定せず (Theory-free)、純粋にデータを観察するだけで政策効果を評価できるという意味で用いられてきた。しかし、ルーカス批判を考慮すると、EBPに必要な「予測のためのエビデンス」を得るためには経済理論が不可欠である、というのが逆説的で興味深い*27。

4. まとめ

他の先進諸国や国際機関とは異なり、我が国では実施した政策の効果が科学的に検証されておらず、エピソードベースで政策立案が行われてきたとの指摘がある。もしこの指摘が正しいのであれば、EBPの推進は既存の政策スキームを改善する一助となるだろう。本稿では、EBPを実行するために必要な政策効果の分析手法として、政策評価法、VAR、そして構造推定を紹介した。こ

れらの手法は想定されているデータ制約や実装難易度、そして得られる結果の性質が異なるため、実施する政策の性質に応じて適切な手法を選択する必要がある。

EBPの推進にあたって、「正しいEBP」とは何かといった議論を行ったり、実施するための枠組みや制度を作ったりすることは重要であるが、そのためには、まず、どのような政策に対してどのような分析・評価法が適切なのかを明らかにするため、様々な事前・事後評価法を試行し、成功と失敗の経験、いわば政策評価のエビデンスを蓄積していくことが肝心であろう。この際、公的統計の充実や利用促進を通じて、行政内部はもとより、外部研究者の研究を促し、その研究結果を積極的に活用するといった取り組みが期待される。

参考文献

- [1] 家子直之・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』, 三菱UFJリサーチ&コンサルティング政策研究レポート。
- [2] 伊藤公一郎 (2016), 『政策の効果をどう測定するか? 海外における「エビデンスに基づく政策」の最新動向』, http://home.uchicago.edu/ito/pdf/RIETI_BBL_2016_1025_Ito_Final.pdf (参照 2017-2-1)。
- [3] 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。
- [4] 黒田昌裕・有本建男 (2010), 『エビデンスに基づく政策形成のための「科学技術イノベーション政策の科学」の構築』, http://scirex.grips.ac.jp/about/download/crds_101028.pdf (参照 2017-2-1)。
- [5] 小林庸平 (2016), 『エビデンスを政策にどう使うか—イギリスの動向と日本への適用事例からの示唆—』, <http://www.rieti.go.jp/jp/events/bbl/16102501.pdf> (参照 2017-2-1)。
- [6] 田中隆一 (2015), 『計量経済学の第一歩——実証分析のススメ (有斐閣ストゥディア)』, 有斐閣。

*26) 政策変更によって不変な構造モデル内のパラメータは構造パラメータ (Structural parameter)、またはディープパラメータ (Deep parameter) と呼ばれ、政策変更によって可変な誘導型パラメータ (Reduced-form parameter) と区別される。ここで注意したいのは、「構造推定だからルーカス批判に抵触しない」のではなく、「パラメータが根源的で政策変更によって不変だから抵触しない」という点である。構造推定か誘導型推定かを問わず、パラメータが根源的でない限り、ルーカス批判は免れない。例えば、先進諸国や国際機関で標準的に用いられているDSGE (動学的確率的一般均衡) モデルでは、潜在的な経済主体の異質性を考慮せず、マクロ経済変数の時系列データに代表的個人モデルを仮定した構造推定を行っている。そのため、代表的個人モデルの構造パラメータが異質性の変化を通じて変化する場合、その予測はルーカス批判に抵触し、信頼出来ないものになる (例えば Geweke (1985), Kirman (1992), Altissimo et al. (2002))。

*27) 本稿では予測に焦点を絞って議論したが、誘導型推定自体に対しても、識別不能性など様々な批判が存在する。例えば Wolpin (2013)、Rust (2014) を参照のこと。

- [7] 田辺智子 (2006), 『エビデンスに基づく教育—アメリカの教育改革と What Works Clearinghouse の動向—』, 日本評価研究 6-1.
- [8] 内閣官房 (2016), 『EBPMのニーズに対応する経済統計の諸課題に関する研究会』, http://www.cas.go.jp/jp/seisaku/ebpm_kenkyukai/index.html (参照2017-2-1).
- [9] 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』, ダイアモンド社.
- [10] 藤田昌久編 (2016), 『日本経済の持続的成長: エビデンスに基づく政策提言』, 東京大学出版会.
- [11] 星野崇宏 (2009), 『調査観察データの統計科学—因果推論・選択バイアス・データ融合 (シリーズ確率と情報の科学)』, 岩波書店.
- [12] 森川正之 (2016) 『「エビデンスに基づく政策」に関するエビデンス』, http://www.rieti.go.jp/jp/columns/a01_0447.html (参照2017-2-1).
- [13] 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, 『進化する経済学の実証分析』, 日本評論社.
- [14] W. ハイゼンベルク (1974), 『部分と全体』, みすず書房.
- [15] Altissimo, Filippo, Stefano Siviero, and Daniele Terlizzese, 2002. "How deep are the deep parameters?" *Annals of Economics and Statistics*, 67-68, 207-226.
- [16] Angrist, Joshua D., 1990. "Lifetime earnings and the Vietnam era draft lottery: evidence from Social Security Administrative Records." *American Economic Review*, 80 (5), 313-336.
- [17] Angrist, Joshua D. and Jörn-Steffen Pischke, 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.
- [18] Brodersen, Kay H., Fabian Gallusser, Jim Koehler, Nicolas Remy, and Steven L. Scott, 2015. "Inferring causal impact using Bayesian structural time-series models." *Annals of Applied Statistics*. 9 (1), 247-274.
- [19] Deaton, Angus, 2010. "Instruments, randomization, and learning about development," *Journal of Economic Literature*, 48 (2), 424-455.
- [20] Friedman, Jerome H., Trevor Hastie, and Rob Tibshirani, 2010. "Regularization paths for generalized Linear models via coordinate descent," *Journal of Statistical Software*, 33 (i01).
- [21] Geweke, John, 1985. "Macroeconometric modeling and the theory of the representative agent," *American Economic Review*, 75 (2), 206-210.
- [22] Head, Megan L., Luke Holman, Rob Lanfear, Andrew T. Kahn, and Michael D. Jennions, 2015. "The extent and consequences of P-hacking in Science," *PLoS Biology*, 13(3).
- [23] Heckman, James J., Robert J. Lalonde, and Jeffrey A. Smith, 1999. "The economics and econometrics of active labor market programs," *Handbook of Labor Economics*, Volume 3A, 1865-2097.
- [24] Imbens, Guido W., and Jefferey M. Wooldridge, 2009. "Recent developments in the econometrics of program evaluation," *Journal of Economic Literature*, 47 (1), 5-86.
- [25] Kirman, Alan P., 1992. "Whom or what does the representative individual represent?," *Journal of Economic Perspectives*, 6 (2), 117-136.
- [26] Lucas, Robert Jr, 1976. "Econometric policy evaluation: a critique," *Carnegie-Rochester Conference Series on Public Policy*, 1 (1), 19-46.
- [27] Nevo, Aviv and Michael Whinston, (2010), Taking the dogma out of Econometrics: structural modeling and credible inference, *Journal of Economic Perspectives*, 24 (2), 69-82.
- [28] Rubin, Donald B., 1978. "Bayesian inference for causal effects: the role of Randomization," *Annals of Statistics*, 6 (1), 34-58.
- [29] Rust, John, 2014. "The limits of inference with theory: a review of Wolpin (2013)," *Journal of Economic Literature*, 52 (3), 820-50.
- [30] Wolpin, Kenneth I., 2013. "The limits of inference without theory," The MIT Press.