

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年 5 月 15 日現在

機関番号：10101

研究種目：挑戦的萌芽研究

研究期間：2010～2011

課題番号：22653105

研究課題名（和文） 大学教育効果をめぐる卒業生調査と大規模社会調査の統合的分析に関する研究

研究課題名（英文） Toward Integrated Analysis of Large Social Survey Data and Data of University Alumni and Alumnae

研究代表者

平澤 和司 (HIRASAWA Kazushi)

北海道大学・大学院文学研究科・准教授

研究者番号：30241285

研究成果の概要（和文）：大学の教育効果を検証する際、大規模な社会調査と、各大学が行っている卒業生調査を統合させて分析することが有益である。しかしその方法論が明らかでなかった。そこで傾向スコア（両調査に共通の変数）を用いてデータ融合を試みたが、現存する卒業生調査データには、そのための共通変数が十分に揃っていないため、難しいことが分かった。あわせて傾向スコアを用いた因果推論に関する問題点を指摘した。

研究成果の概要（英文）：It is useful to merge large social survey data with data of alumni and alumnae of a university to verify effects of university education on occupational attainment. So I try to merge them by using propensity score, but it is difficult to do so because the later do not contain enough covariate variables. We have to think propensity score carefully to apply it to social phenomena.

交付決定額

（金額単位：円）

| | 直接経費 | 間接経費 | 合計 |
|--------|-----------|---------|-----------|
| 2010年度 | 600,000 | 0 | 600,000 |
| 2011年度 | 500,000 | 150,000 | 650,000 |
| 年度 | | | |
| 年度 | | | |
| 年度 | | | |
| 総計 | 1,100,000 | 150,000 | 1,250,000 |

研究分野：社会学

科研費の分科・細目：教育学・教育社会学

キーワード：卒業生調査、社会調査、大学教育、不完全データ、傾向スコア

1. 研究開始当初の背景

近年、多くの大学が卒業生を対象にした調査を行っている。その背景には、大学の生き残り競争が激しさを増すなかで、大学の教育効果を定性的・定量的に把握し、市民への説明責任や受験生の確保に役立てたいという意図が共通してあると考えられる（吉本2007）。

これは調査の実践的な意義であるが、学術的にはそれぞれの大学に固有の教育効果を

定量的にどのように推定するというかという問題がある。これまでいわゆる官庁統計（賃金センサス、就業構造基本調査など）や大規模社会調査（SSM 社会階層と社会移動に関する全国調査、JGSS 日本版総合的社会調査など）によって検証されてきた大学教育の効果（たとえば所得の増加、威信の高い職業へ到達率の増加など）は、大学を一括りにしたときの、他の学歴の者（通常は高卒者）に対する平均的な効果である。そのため個別

の大学の効果はわからない。それに対して、各大学の卒業生調査では、みずからの大学の卒業生しかデータに含まれていないため、当該卒業生間の分散の説明にとどまり、他の学歴や他の大学の卒業生との教育効果の比較は直接できない。

この問題を解決する方法は二つ考えられる。ひとつは大規模調査の標本の大きさを大きくして、そこに含まれる各大学の卒業生数を増やすという方法である。この場合、大学ごとの卒業生調査はかならずしも行う必要がない。しかし全国の大学数、卒業生の出生コーホートの広がり、大規模調査の実施年の間隔を考えると、このやり方は現実的ではない。もうひとつは、大学別の卒業生調査と大規模社会調査を統合して分析する方法である。具体的には、両データの比較や併合しての分析である。こちらのほうが実現が容易であろう。

しかし研究開始当初は、大規模社会調査と、各大学の卒業生調査の情報をうまく共有できれば有益な知見が得られる可能性があるにもかかわらず、その方法論が確立されていないために、両者が完全に分離されたまま分析されていた。

2. 研究の目的

そこで本研究では、大学の教育効果に関して、大規模社会調査データと各大学の卒業生調査データを統合して分析するには、どうすればよいのかについて、その考え方と条件の提案を目的とする。

具体的には、

(1) まず大学教育の成果や効果を測定する変数、および大規模社会調査と大学卒業生調査の双方に含まれている変数を特定する。後者の共通変数は、両者を統合して分析するときの鍵となる。

(2) 大規模社会調査と各卒業生調査という母集団と抽出確率がまったく異なるデータセットに対して、どうすれば両データセットをひとつのデータセットとして、大学教育効果を分析できるのか、(1)の共通変数を中心に、その具体的な方法論を模索する。

3. 研究の方法

(1) 研究目的(1)に関して

まず大学教育の成果や効果が、具体的にどのような変数で測定されているのかに関して、これまでに卒業生調査を行ったいくつかの大学の報告書から調べる。ただしそれだけでは、なぜそうした変数を盛り込んだかわからないので、必要に応じていくつかの大学

の調査担当者から、詳細な聞き取り調査を行う。

つぎにそれらの変数が、大規模社会調査に含まれているかを確認する。その際、大学教育効果を測定する変数だけではなく、どういった学生が大学に入学しているかが分かる変数が、卒業生調査と大規模社会調査の双方に含まれているかどうかを確認することが重要である。

(2) 研究目的(2)に関して

なぜなら、過去 20 年ほどの間に、アメリカの社会学や経済学の分野で、因果効果の大きさに関してパラダイム・シフトがおきているからである。従来は、大規模社会調査データを用いて、独立変数 X (例、大学教育を受けたかどうか) が、大学教育の効果とされる従属変数 Y (高卒者に対する賃金の増加) に与える影響を知るには、統制変数 Z (年齢や性別など) を独立変数に加えた (何らかのリンク関数による) 重回帰分析を行えばよく、結果的に独立変数 X の偏回帰係数に効果の大きさが表現されると考えられてきた。特定の傾向の (たとえば高校までの学校の成績が高い) 者が大学教育を受けやすい、という選択バイアスを除去するために用いられるヘックマンの 2 段階推定法も、広義ではこの範疇に入る。

しかしこれらの方法は前提となる仮定が強く (たとえば共変量が従属変数に与える効果が通常は線形関係に特定されている)、かならずしも頑健でないことが指摘されている (星野 2009)。その代わりに提案されているのが、傾向スコアを用いる方法である (Rosenbaum and Rubin 1983)。これは実験計画法における無作為割り付けを、社会調査に応用しようとするものである。もし対象者を高校卒業時点で、最終学歴が高卒者と大卒者のいずれかの群に無作為に割り付けることができれば、両群間では、共変量 Z (性別や生まれつきの能力など高卒以前の特性) の分布が等しくなるはずである。それにもかかわらず両群間で卒業後の賃金 (Y) に差があるとすれば、それは大学教育を受けたかどうか (X) によって生じたといえるだろう。そこで、その差を因果効果の大きさとみなすことができる。

もちろん現実社会では、こうした無作為割り付けなど倫理的に許されるはずもない。そこで、無作為割り付けによる実験ではなく、対象者が無作為抽出された通常の実験データにおいて、複数の変数からなる共変量 (Z) を一元化することが提案されている。

それが傾向スコアで、算出された各人の傾向スコアは、ここでは、各人が大学教育を受ける傾向（受けやすさ）を表していると考えられる。

この傾向スコアが同じか近い者のなかで、実際の高卒者と大卒者の賃金（Y）を比較すれば、それが因果効果の平均的な大きさになる。こうすることによって、もともと何らかの能力の高い者が大学に行きやすいといった選択バイアスを除去できたと見なすからである。この方法が従来の重回帰分析とまったく異なるのは、すでに述べたように、共変量と従属変数との関数を特定する必要がない点である。

以上の考え方は、本質的には反実仮想モデル（Holland 1986）に依拠して選択バイアスを除去することなので、こうした因果推論だけではなく、不完全データの分析にも容易に応用できる（星野 2009）。不完全データとは、たとえばインターネット調査など、対象者が無作為抽出されていないデータを指す。それと、無作為抽出された大規模社会調査の双方に含まれる共通の変数が上記の共変量に相当する。それをを用いて有意抽出されたデータを、あたかも無作為抽出されたデータのように補正して、正しい因果効果の大きさが推計できるとされる（星野 2009）。この考え方を、一種の不完全データである卒業生調査に応用しようというのが、本研究での当初の基本的な戦略である。

4. 研究成果

(1) 大学教育の成果に関する質問項目

CiNi で「大学、卒業生、調査」を検索語として検索したところ、研究を開始した時点で約 600 件、「大学卒業生、調査」では約 200 件がヒットした。それらのなかから大学が調査主体で、調査票か調査項目が分かり、かつ調査報告書が容易に入手可能な調査は約 30 件あった。

各調査によって目的が異なるので、当然さまざま事項が調査されている。そのなかから教育効果と思われる項目を取り出すと、やや恣意的ではあるが、職業に関する客観的な項目（大学卒業後はじめてついた仕事と現在の仕事の職種、企業規模、従業上の地位、および現在の個人年収など）と、職業に関する主観的な項目（仕事の進め方や考え方、仕事から得られる満足感、大学で学んだことと仕事に関連しているか、大学で学んだことが仕事に役に立っているかについての考えなど）に分けられる。

(2) 比較すべき対象者と利用可能なデータ

これらの項目をひとまず大学教育の成果と見なした場合、まず検討すべきなのは、当該大学の結果を、高卒者と比較することである。これは賃金であれば、比較的容易に行える。官庁統計には、性別・学歴別・年齢層別の賃金が集計されているからである。

他方、聞き取りを行った多くの大学では、他の大学の卒業者と比較して自校がどの位置にあるのかを知りたがっていた。しかしこれを確かめることは、意外に難しい。なぜなら、4 年制大学への進学率が 50% に近い今日、大学進学者の内実は一様ではなく、大卒者を一括りにした賃金と比較しても、情報としては相当に粗いからである。関心の焦点は高卒者との比較から、他の高等教育修了者との比較へ移ったとあってよい。

したがって、何らかの基準によって大学を分類したうえで、その大学群ごとの賃金を知りたいところである。ただし大学をどう分類するのが適切かは定説がないので、平沢（2010、2011）では設置者と創立時期によって、国公立の銘柄大学、それ以外の国公立大学、戦前設立の私立大学および戦後設立の私立大学の 4 つに暫定的に分けてみた。

こうした大学分類は、官庁統計では当然なされていないので、以下の分析では大規模社会調査を利用することになる。もっとも、管見の限り、①個票データが公開されていて（つまり 2 次利用が可能で）、②全国規模で無作為抽出がなされており、③幅広い年齢層を含んでいるうえに、④回収率が一定以上あり、⑤大学分類が可能な大規模社会調査となると、SSM 調査、JGSS 調査、および JLPS 調査（働き方とライフスタイルの変化に関する全国調査）しか見あたらない。

さらに 4(1) に掲げた大学教育の成果に関する変数をほぼすべて含んでいて、かつ詳細な大学分類が可能なのは、JLPS 調査だけである。ただし同調査は、①対象者が 20~40 歳（2007 年）に限られているうえに、②パネル調査として実施されているために、第 1 波調査時に長期間の協力を得られる者のみが回答することが影響してか、回収率が約 36% にとどまっている。

よって厳密には、大規模社会調査といっても、上記の条件をすべて満たしたデータは皆無ということになる。条件をゆるめても、きわめて限定されることをまず認識する必要がある。

(3) 卒業生調査と大規模調査に共通の変数

つぎに無作為抽出された大規模社会調査

と、通常は有意抽出の各大学の卒業生調査とをデータ融合させるには、双方に共通の変数が、経験的には「20 から 30 程度」(星野 2009: 185) 必要とされる。たとえば星野・前田 (2006) が JGSS 調査とインターネット調査を融合させたときは、性・年齢・地域ブロック・都市規模・学歴・父学歴・母学歴・職業分類・世帯収入・住居形態・新聞を読む頻度・喫煙・飲酒・先週の状況・先週仕事をしてたかどうか・所属階層の自己評定・同居人数・仕事でメールをするか、の 18 変数を用いている。

JGSS のほか、他の 2 つの大規模社会調査 (SMM と JLPS) も、これらの変数の多くを含んでいるので、融合に基本的に問題はない。問題なのは、各大学の卒業生調査のほうである。収集した卒業生調査の報告書から判断する限り、これらの変数をすべて含んでいた調査は一つもなかったからである。

もちろん共通の変数は、以上の通りでなければならない必然性はない。最低でも 20 程度の共通変数があればよい。ただし共通変数は、原則として独立変数と従属変数とに影響を与えるいわゆる交絡変数でなければならない。換言すれば、独立変数と従属変数を媒介する中間変数は望ましくない (星野 2009: 118-122)。

したがって、データ融合を目指す場合には、卒業生調査を設計する段階で、何を従属変数と独立変数にするかを決めたいうえで、適切な共通変数を選択し調査票に盛りこまなければならないということである。

(4) 大学教育の効果と傾向スコア

このように現状では、大規模社会調査と卒業生調査を、共通変数によってデータ融合させるのは、多くの困難が伴うことがわかった。さらに傾向スコアを用いた補正法を、社会調査データに適用する際には、留意すべき点があるように思われる。

大学の教育効果を検証するには、大学に進学したこと (学歴)、あるいはいずれの大学 (群) に進学したこと (学校歴) が独立変数となり、その成果をあらゆる諸変数の分散を説明する、というのが基本的な分析枠組みとなる。つまり学歴や学校歴が、傾向スコアで補正されるべき割り付け変数 (独立変数) である。

このときこの学 (校) 歴の効果は、高卒時にいずれの学歴 (または大学群) に入学したかによって決まるのか、それともいずれの学歴 (大学群) で教育を受け卒業したかによって決まるのか、を区別して考える必要がある。

なぜなら傾向スコアで補正できるのは、前者だけであり、いずれかの学歴 (大学群) に、擬似的に割り付けられた後の効果は両群で等しいことが仮定されているからである。

しかし直感的には、双方の効果があると想定するのが自然であろう。つまりもともと何らかの能力の高い者が大学に行きやすいためだけでなく、特定の大学 (群) を卒業したことによって、卒業後の賃金が異なってくるのであり、しかも入学後の効果は、同年齢層のうちどのくらいの人々が各学歴や各大学から卒業したかによって異なると思われる。こうしてスクリーニング論と人的資本論のいずれが正しいかという、解決困難な古典的な問題に立ち戻ってしまうことになる。

要するに、傾向スコアは実験における無作為割り付けを、社会調査データ上で擬似的に行うためのツールであるが、それを大学教育効果といった長い時間を伴う複雑な社会現象の分析に、そのまま応用してよいのだろうか。慎重に議論を深化させることが肝要であるように思われる。

文献

- 平沢和司 (2010) 「大卒就職機会に関する諸仮説の検討」 荻谷剛彦・本田由紀編『大卒就職の社会学』東京大学出版会、61-85 頁。
- 平沢和司 (2011) 「大学の学校歴を加味した教育・職業達成分析」 石田浩・近藤博之・中尾啓子編『現代の階層社会 2 階層と移動の現在』東京大学出版会、155-170 頁。
- Holland, P. W. (1986) "Statistics and Causal Inference," *Journal of the American Statistical Association*, vol. 81, 945-960.
- 星野崇宏 (2009) 『調査観察データの統計科学』岩波書店。
- 星野崇宏・前田忠彦 (2006) 「傾向スコアを用いた補正法の有意抽出による標本調査への応用と共変量の選択法の提案」『統計数理』第 54 巻第 1 号、191-206 頁。
- Rosenbaum, P. R. and D. B. Rubin (1983) "The Central Role of the Propensity Score in Observational Studies for Casual Effects," *Biometrika*, vol. 70, 41-55.
- 吉本圭一 (2007) 「卒業生を通じた「教育効果」の点検・評価方法の研究」『大学評価・学位研究』第 5 号、25-107 頁。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)
なし

6. 研究組織

(1) 研究代表者

平澤 和司 (HIRASAWA Kazushi)

北海道大学・大学院文学研究科・准教授

研究者番号：30241285

(2) 研究分担者 なし

(3) 連携研究者 なし