

教員・学校評価におけるSGPモデルの方法的検討

学校開発政策コース 橋野晶寛

A Methodological Examination of SGP Models for Teacher/School Evaluation

Akihiro HASHINO

Abstract

It is well known among researchers of education policy that teacher/school evaluation systems in the U.S. are based on 'student growth measure'. However, technical details of these evaluation systems are not widely known. The methodological basis for calculating student growth is the estimation of the status conditioned on prior test scores and covariates via regression analysis. In this study, we examine one such statistical models namely the Student Growth Percentile (SGP) model using nonlinear quantile regression, and its significance in the context of educational accountability systems. We address issues relating the statistical properties of the SGP model by comparing it to other familiar model Value-Added Models, with an aim to improve the existing SGP model. Based on simulations, our results indicate that the augmented SGP model are not inferior to VAMs on the ground that it can successfully control student level covariates and school level unobservable.

目次

1. 問題の所在
2. SGPモデルの統計的性質の検討
 - A. 分位点回帰モデルの応用としてのSGPモデル
 - B. 付加価値モデルとの比較
 1. 成績変数データの性質
 2. 観測可能な要因・不可能な要因の統制
3. SGPモデルの拡張
 - A. 解決すべき問題
 - B. 代替SGPモデルの検討
 1. 線形固定効果モデル
 2. 部分線形変量効果モデル
 - C. シミュレーションデータによる比較
4. 考察

1. 問題の所在

本稿は、アメリカの教員評価・学校評価の統計モデルの1つである「成長パーセンタイル」(Student Growth Percentiles: SGP) モデルの統計的性質を検討するものであり、その方法的意義とアカウンタビリティ政策の文脈における意義、教育政策過程における研究—社会間関係への示唆の解明を企図している。

教育行政・教育経営における評価の問題は、研究者共同体内部で閉じた学術的研究、あるいは、実務家間で完結した行政実務ではなく、双方の間での相互交渉を伴う。すなわち、教員評価・学校評価自体は極めて実務的な行政行為であるが、その方法に関わる技術の開発・検討、実装・支援には研究者共同体の関与が必要とされる。

1990年代以降、アメリカの各州で、大規模テストに基づいた教員評価・学校評価システムの適用・開発が進められてきたことは日本の教育政策研究者にもよく知られている。そうした初等中等教育におけるアカウンタビリティ政策はこの20数年の間に観察された潮流であるが、とりわけ、NCLB法およびオバマ政権期における競争的連邦補助金Race to the Topがそのトレンドを後押しすることとなった。2015年の初等中等教育法第7次改正において連邦の関与は後退したものの、依然として多くの州が成長指標¹⁾を用いた評価システムを維持している。

学術的な文脈での議論においては、そうした教員・学校評価のハイステークスな局面での運用には慎重論が大勢を占めるが、評価技術によって産出される評価情報自体はローステークスな用途においては一定の有用性がある(Muller 2010=2019)。社会科学全般における因果推論の文脈でもその手法に関する議論は少な

からず関心を集めている。

標準化テストを用いた評価手法としては、付加価値モデル (Value-Added Model: VAM) がよく知られている。VAMは、パネル・クラスターデータでの線形回帰モデルによる分析手法を教員評価・学校評価モデルに転用したものであり²⁾、教育統計学、教育経済学、計量経済学の各分野の研究者によって無数の方法的検討が加えられてきた。また、応用研究者の間でも広く知られており、付加価値モデルは、評価手法への適用以前にも古くから教育政策に関する実証研究の局面で用いられてきたという経緯があり、比較的浸透した方法と言えよう (Bryk&Weisberg1976; Hasnuhek 1972; Murnane 1975)。

VAM以外の大規模テストのデータに基づく評価手法としては、成績推移表 (Transition Table)³⁾、SGPモデルがあり、後者は、Betebenner (2009) によって提唱された比較的新しい評価手法であり、コロラド州を皮切りとして用いられてきた⁴⁾。評価システムとしてこのSGPモデルを採用している州数は2015年時点で19州であり、VAMの15州を上回っている (Kurtz 2018)。SGPモデルもVAM同様回帰モデルを応用した手法であるが、教育政策研究者の間でさほど関心が払われてきたとは言い難い。それはVAMと同じくパネルデータの回帰分析の手法に依拠しながらも、分位点回帰モデルという政策研究者に馴染みの薄い方法を基盤としていることに拠っている。しかし、一方で、SGPモデルによって得られる情報自体は、後述のように実務家にとって極めて理解しやすいものであり、透明性の点においてSGPモデルが教員評価・学校評価に採用されてきたものと思われる。

以下では、SGPモデルについて、評価手法としての意義と可能性の検討を行う。次節ではSGPモデルについて、分位点回帰の応用としての統計モデルに立ち戻り、その性質・推定法、VAMとの異同を整理する。第3節では、プロトタイプのSGPモデルの拡張として、固定効果分位点回帰モデル、部分線形分位点回帰モデルに基づく手法を示し、シミュレーションデータにおける比較を通じて既存モデルとの対応関係を明らかにする。第4節では、それらの議論をふまえて、SGPモデルの統計モデルとしての意義と可能性に加えて、研究—社会間関係、教育政策過程の文脈における示唆を述べる。

2. SGPモデルの統計的性質の検討

A. 分位点回帰モデルの応用としてのSGPモデル

SGPモデルは、Betebenner (2009) によって提起された教員評価・学校評価のための統計モデルであり、その技術的中核はノンパラメトリック分位点回帰モデルに基づいている⁵⁾。児童・生徒個人のテストスコアを被説明変数とする回帰モデルという点でVAMと性格を共有しているが、Betebenner (2009, 2011) によるオリジナルのSGPモデルでは学校・教員の個別効果の因果推論よりも記述的側面が強調されている。教員評価を用途としたSGPモデルは以下のように示すことができる。

教員*i*のクラスに属す生徒*j*の時点*t*のテストスコアを y_{ijt} とする。最も単純なケースとして、それが過去の時点*t-1*のテストスコア $y_{ij,t-1}$ の関数となる回帰モデルを考える。通常回帰モデルでは、条件付き期待値関数の推定に関心があるが、分位点回帰モデルでは条件付き分位点関数を推定する。変数*y*の τ 分位点を $Q_y(\tau)$ として表す時に、 $y_{ij,t-1}$ に条件づけた τ 分位点 $Q_y(\tau|y_{ij,t-1})$ を式(1)のように表す⁶⁾。

$$Q_y(\tau|y_{ij,t-1}) = f_\tau(y_{ij,t-1}) \quad (1)$$

関数 $f_\tau(\cdot)$ は、 τ の値によって異なることを意味している。最も単純な関数形は、式(2)のような線形関数によるものである。

$$f_\tau(y_{ij,t-1}) = \beta_{0\tau} + \beta_{1\tau}y_{ij,t-1} \quad (2)$$

SGPでは、関数形にこのような先験的な仮定を置かず、スプライン関数を用いて $f_\tau(y_{ij,t-1})$ をノンパラメトリックに推定することを考える。Bスプライン基底関数を用いて、式(3)のように近似することができる⁷⁾。

$$f_\tau(y_{ij,t-1}) = \sum_h \beta_{h\tau} B_{h\tau}(y_{ij,t-1}) \quad (3)$$

条件付き分位点関数を求めるには、式(4)を最適化するパラメータを推定する。

$$\arg \min_{\theta} \sum_{i=1}^N \rho_\tau(y_{ijt} - f(y_{ij,t-1}; \theta))$$

$$\rho_{\tau}(u) = u \cdot (\tau - I(u < 0)) \tag{4}$$

上記の最適化は線形計画法によって実行可能である。また、分位点回帰モデルは、後述のように、非対称ラプラス分布に従う誤差項をもつ回帰モデルとして表すことができ、代替的な方法でパラメータ推定を行うこともできる。

こうして複数の τ について条件付き分位点関数を推定し、 y_{ijt} が式(5)の関係を満たすならば、観測値 y_{ijt} は、 y_{ijt-1} に条件づけた時に、 $100\tau'$ パーセンタイル点よりも大きく、 $100\tau''$ パーセンタイル点よりも小さいという関係が成立している。

$$f_{\tau'}(y_{ij,t-1}) < y_{ijt} < f_{\tau''}(y_{ij,t-1}) \tag{5}$$

この $100(\tau' + \tau'')/2$ の値を各生徒の「成長指標」として解釈する。直観的にこの成長指標が意味するのは、時点 $t-1$ において同じ成績であった集団の中で時点 t の各生徒の成績が下位から何パーセントの位置にあるか、ということである。視覚的には図1のように表現することができる。

図1-1は横軸に時点 $t-1$ における成績、縦軸に時点 t における成績を示した散布図である。図1-2はこの散布図中に9本の条件付き τ 分位点関数 $Q_{\tau}(\tau | y_{t-1}) = f_{\tau}(y_{t-1})$ の曲線を加えたものである(下から順番に $f_{0.1}$

$(y_{t-1}), f_{0.2}(y_{t-1}), \dots, f_{0.9}(y_{t-1})$)。図中の点*は、1期前の成績が70点で、今期の成績が66点である生徒のデータを示している。この点は $Q_{\tau}(0.4|y_{t-1})$ と $Q_{\tau}(0.5|y_{t-1})$ の曲線に挟まれており、1期前に同様の成績だった者との比較において、40-50パーセンタイルの範囲にあることを意味する。そして、この τ の間隔をより小さくとれば、より精度の高い成長指標を求めることができる。例えば、 τ が0.005, 0.015, ..., 0.995というように間隔が0.01刻みで設定されるならば、各生徒の成長指標は、0から100までの整数値で得られる。

こうして各生徒について求められた成長指標を所属学級もしくは学校ごとに平均値・中央値によって要約した値が、評価対象となる各教員・学校の寄与ないし業績に関する情報となる。ここで示した例は、過去の1時点の成績に条件づけたモデルであるが、複数時点の過去の成績に条件づけたモデルも想定可能である。

SGPモデルは、統計モデルとしては、通常のVAMが基づく回帰モデルよりも馴染みの薄い分位点回帰モデルを用いたものであるが、産出する情報——過去の成績に照らした今期の相対的な成績——は理解が容易であり、おそらくこの透明性における利点ゆえに、VAMよりも採用州が多くなっていると思われる。また、SGPモデルは、大規模テストスコアについて、通時的にテストスコアが比較可能なるようにテストを計画し、項目反応理論を適用する垂直尺度化の手続きを必要としない点でも実務上の大きな利点がある。尚、

図1-1 成績の散布図

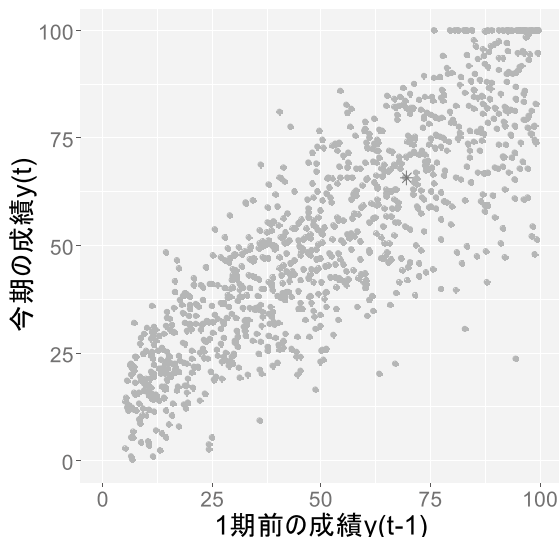
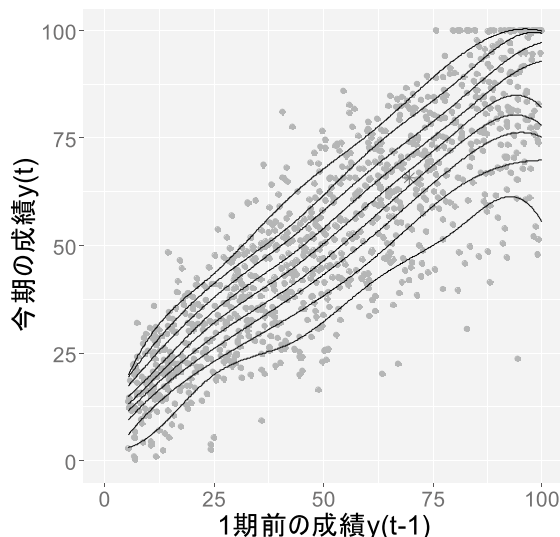


図1-2 条件付き分位点関数



SGP モデルの開発者である Betebenner は SGP モデルを記述に限定した用途のみを想定している。すなわち、過去の成績以外の共変量を加えたモデルを想定していない。この点は、次に述べるように、SGP の適用可能性をめぐる重要な論点となる。

B. 付加価値モデルとの比較

SGP の特性について、同じく成長指標として言及される機会の多い VAM との異同を明らかにしておくことは有益である。SGP と VAM との比較については、少ないながら既に研究がいくつかなされており、以下簡潔に論点を整理しておく。

まず、SGP と VAM との共通点としては、共に回帰分析を用いた conditional status に関する指標であると要約することができる。VAM や SGP は、しばしば成長指標に基づくものとして言及されることが多いが、その成長指標の定義は必ずしも明確ではない。筆者の理解では、「成長」という語には狭義と広義の意味が存在する。広義の「成長」は、過去の成績に条件つけた現在の成績の相対的位置を意味するのに対して、狭義の「成長」はさらにその大きさが増分の間隔尺度として解釈できることを意味している。SGP が依拠する指標は、広義の意味では成長指標である一方、狭義の意味では成長指標ではない。つまり、特定の児童・生徒が「どれだけ伸びたか」という点に関して情報を持った評価指標ではない。VAM が依拠するものには狭義の成長指標であるものとそうでないものが存在するゆえに、SGP は VAM と、回帰分析を用いた広義の意味での成長指標の関数という性質を共有していると言える。

狭義の意味で成長指標に基づくものではないという点を肯定的に捉えるならば、それは、テストスコアなどの成績変数について、スコアを通時的に比較可能にするための垂直的尺度化が必ずしも必要でなく、適用できるデータの範囲が広いことを意味している⁸⁾。教員・学校の評価のために相対的なランキングを得ることのみに目的があるのならば、この点は実用性の面で大きな利点をなす。

一方で、SGP と VAM との間で生じうる相違点としては、1) 成績変数の外れ値・歪みに対する頑健性、2) 尺度変換に対する頑健性、3) 「成長指標」の精度、4) 観測可能な共変量による条件付けの有無、4) 欠落変数による内生性への対処、が指摘できる。1), 2) は回帰モデルの被説明変数である成績変数のデータの性質に関するもの、3), 4) は説明変数の性質に関

わるものである。

1. 成績変数データの性質

成績変数の外れ値・歪みがもたらす影響については、Castellano&Ho (2013) において、実データとシミュレーションデータによる考察が行われている。VAM では回帰関数が成績の条件付き期待値を与えるのに対し、SGP が基づく分位点回帰モデルでは回帰関数が成績の条件付き分位点を与える。Castellano&Ho (2013) では前者の残差をパーセンタイル変換した PRR (percentile rank of a residual) と SGP との比較を行った結果⁹⁾、PRR による成長指標と SGP による成長指標との間に非常に高い相関を見出している。どちらも不偏推定量となっているが、複数年の成績データ ($Y_{it}, Y_{i,t-1}, Y_{i,t-2}, \dots$) が多変量正規分布に従うデータでは PRR の方が効率的となっている。一方で、多変量正規分布に従わず外れ値のあるデータ、非線形の尺度変換を施した成績データでは、双方の間に相違が生じ、SGP の方が頑健であるという結果を得ている¹⁰⁾。こうした成績データの分布の歪みによる VAM と SGP による相違は Kurtz (2018) によっても同様に指摘されている¹¹⁾。

VAM と SGP を教員評価・学校評価に用いるには、必然的に成長指標の集計方法——平均値によって要約するか、中央値によって要約するか——の問題が付随する。VAM では、評価の対象となる教員・学校効果について、2 種類の異なる方法がある。1 つは 1 段階 VAM (one-step VAM) であり、固定効果モデル・変量効果モデルで回帰係数とともに教員効果・学校効果を一度の手続きで推定する方法である。もう 1 つは 2 段階 VAM (two-step VAM) であり、第 1 段階で教員効果・学校効果を含まない回帰モデルを推定し、第 2 段階でその残差を学級・学校ダミー変数群に回帰させる——もしくは、学級・学校レベルで平均値を求める——方法である¹²⁾。いずれも平均値による集計を想定している点では共通している。一方 SGP では、パーセンタイルランクとして得られた児童・生徒個人レベルの成長指標を中央値で要約することが多い。Castellano&Ho (2015) では、こうした通常の VAM、SGP の手続きから、改めて集計方法の問題を切り離れた上で SGP と VAM の比較を行った。そこでは、1) 平均ベースの残差指標 (固定効果モデル、変量効果モデル、回帰残差平均)、2) 平均ベースのパーセンタイルランク指標 (SGP の集団平均、PRR の集団平均)、3) 中央値ベースのパーセンタイルランク指標 (SGP の集団中央値、

PRRの集団中央値)に類型化され、同一類型内での強い相関と類型間での大きな相違を見出している。この比較に基づけば、従来のSGP/VAMの間の相違の一部は、集計方法の相違に帰せられることになる。

2. 観測可能な要因・不可能な要因の統制

SGPが提唱されてから比較の日が浅い時期に行われたこうしたCastellano&Ho (2013, 2015)の知見は興味深いものだが、VAMを巡る論争において中核となっている因果推論との接点は希薄である。因果推論の文脈の中でSGPとVAMとを比較した分析の多くは経済学研究者によって行われている。焦点となるのは、SGPと観測可能な共変量を条件づけたVAMとの相違、SGPと非時変的な観測不能な要因を固定効果で統制できるVAMとの相違である。SGPは、通常、過去の成績のみを共変量としたノンパラメトリック分位点回帰モデルによって推定がなされる。技術的には過去の成績の関数のノンパラメトリックな部分を維持しつつ、過去の成績以外の共変量——児童・生徒の社会経済的背景や学校レベル変数——を回帰モデルに含めることは可能である。しかし、管見の限り、そのようなモデルによるSGPは用いられていない。また、VAMにおいて最もよく用いられる線形固定効果モデルは、変数の級内平均を観測値から差し引くというdemean変換した変数によっており、非時変的要因の統制が容易であるが、分位点回帰モデルにおいては、このような簡便な手続きがないため、因果推論の観点からSGPと固定効果VAMとの間に相違が生じうる。その際、実務的に重要な問題となるのは、それらによる教員・学校評価は元々不利な指導環境——社会経済的に不利な児童・生徒、英語を母国語としない児童・生徒が多い等——に置かれた教員・学校に対して厳しい評価指標となるのか否かという中立性に関わる問題である。

Goldhaber *et al.* (2015) は、median SGPと共変量の異なる様々な固定効果VAMとの比較を実データにおいて行っている。そこではSGPとVAMとの相関の高さが指摘されているものの、median SGPによる教員評価は、不利な学校環境——無料給食の児童割合、非白人児童割合の多さ——に属す教員に辛い評価となりやすく、学校レベルの固定効果を含むVAMが最も中立的な評価になることを示している。同様の知見はEhlert *et al.* (2016), Guarino *et al.* (2015), Walsh&Isenberg (2015) においても示されている。

Ehlert *et al.* (2016) は、median SGP, one-step VAM (固定効果VAM), two-step VAM (残差平均VAM) との

間で比較を行っており、3つのモデルでの成果指標間での相関は0.8を超え、高い水準にあるものの、SGPは環境の有利/不利に対して中立的ではなく、two-step VAMではほぼ中立的な評価になることを報告している¹³⁾。

Guarino *et al.* (2015) も同様に、median SGP, 固定効果VAM, 平均残差VAMとの間で比較を包括的に試みており、学級編制および教員割当が非ランダムとなるケースを想定したシミュレーションデータでの分析が行われている。やはり、SGPとVAMによる成果指標は相関が高いものの、学級編制、教員割当が非ランダムなケースでは、SGPによる成果指標推定値と成果指標の真値との相関は低くなり、カテゴリカルな処遇反映にその成果情報を用いた場合、誤分類が発生する可能性が高くなる。外れ値の多い一部のケースではSGPに分があるものの、多くのケースではSGPは固定効果VAMに劣るという結論が得られている¹⁴⁾。

このように既存研究におけるSGPとVAMの比較では、必ずしもSGPに大きな利点を見出していないが、そうした結論は分位点回帰モデルの統計学的性質が未解明であること、もしくは、制約的なモデルのみを扱っていることに帰せられる。しかし、現時点の水準に照らしてもSGPモデルはいくつかの点で拡張の可能性はあるように思われる。

3. SGPモデルの拡張

A. 解決すべき問題

SGPモデルの教員・学校評価への適用が問題となるのは、因果推論の文脈においてである。Betebenner (2009) はあくまでもSGPの用途を記述に限定しているが、実際にはそうした記述的用途だけでは評価指標としての実質的な意義が小さく、あるいは、こうした禁欲的な用途を超えて「乱用」される可能性もある。前節での議論をふまえるならば、因果推論において焦点となるのは、次の2点である。

第1は、既存のSGPモデルでは、現期の成績スコアを過去の成績スコアのみによって条件づけており、成績スコアに影響する観測可能な共変量を統制していない点である。しかし、この点については、セミパラメトリックモデルを採用することで技術的には解決可能である。過去の成績以外の共変量Xを、式(6)のような部分線形モデル、または、式(7)のような単一インデックスモデルにおいて左辺に加えることが考えられる。

$$Q_y(\tau|y_{i,t-1}, \mathbf{X}_{it}) = f_\tau(y_{i,t-1}) + \mathbf{X}_{it}\beta_\tau \quad (6)$$

$$Q_y(\tau|y_{i,t-1}, \mathbf{X}_{it}) = g_\tau(\mathbf{Z}_{it}\theta_\tau) \\ \mathbf{Z}_{it} = [y_{i,t-1} \ \mathbf{X}_{it}] \quad (7)$$

第2は、SGPモデルでは、非観測要因を統制することが難しい点である。VAMで用いられているパネル線形固定効果モデルでは非時変的な個体効果の除去が容易であるゆえに大きな利点があるが¹⁵⁾、分位点回帰モデルではそのような簡便な個体効果の除去法がない。パネル分位点回帰モデルにおいても固定効果モデルがいくつかの理論研究によって考案されているが (Arellano&Bonhomme 2016)、応用研究者が用いるには難があり、また教員・学校評価の文脈での大規模データには適合的でないと思われる方法もある。

そして、これらに加えてSGPモデルには、条件付き分位点関数の交差に関する制約がある。SGPモデルでは様々な τ の値の下で条件付き分位点関数を推定することになるが、 $\tau' < \tau''$ ならば、 $Q_y(\tau'|y_{i,t-1}, \mathbf{X}_{it}) < Q_y(\tau''|y_{i,t-1}, \mathbf{X}_{it})$ を必ず満たす必要がある。すなわち条件付き分位点関数が交差しないように制約を課す必要がある。サンプルサイズが大きく、どの説明変数の区間においてもデータが十分に存在するならば、明示的に制約を課さずとも交差が生じない可能性が高いが、そうでない場合は、何らかの制約を課す必要がある。He (1997) は誤差のスケールパラメータが説明変数に依存するモデルによって交差の問題の解決を図っている。

これらの問題をセミパラメトリックモデルの範囲内で全て同時に解決することは筆者の能力を超えるが、第1の過去の成績以外の共変量の導入と第3の非交差制約は比較的容易に解決可能である。以下では、試論的に従来のSGPモデルの代替モデルを提示し、シミュレーションデータに基づくVAMとの比較を試みる。

B. 代替SGPモデルの検討

1. 線形固定効果モデル

分位点回帰モデルにおける固定効果モデルは、計量経済学で推定法の理論研究が試みられており、決定的と言えるものはないが、積率法に基づくMachado&Silva (2019) は推定が比較的容易であり、大規模データにも十分適用可能であると思われる。また、条件付き

分位点関数の交差の問題についても対応している点でSGPモデルに向いている。

Machado&Silva (2019) は線形モデルであれば、実行が容易である。SGPモデルでは、過去の成績スコアをセミパラメトリックな非線形関数 $f(y_{i,t-1})$ によって統制するが、この部分のモデル化を線形関数で表現できるならば、Machado&Silva (2019) の固定効果モデルは有力な方法となる。過去の成績以外の共変量および集団レベルの固定効果で統制できるという点では、通常のVAMに相対的に劣る点はない。

具体的には、He (1997) と同様に、式(8)のような誤差項のスケールを説明変数の関数とする location-scaleモデルによってパラメータ推定し、その推定値を分位点回帰モデルにおけるパラメータに変換するという手続きをとる。固定効果に関わる集団 (学級もしくは学校) に関する添え字 i とすると固定効果 location-scaleモデルは

$$y_{it} = \alpha_i + \mathbf{Z}_{it}\beta + \sigma(\delta_i + \mathbf{Z}_{it}\gamma)\epsilon_{it}$$

$$E[\epsilon] = 0$$

$$E[|\epsilon|] = 1 \quad (8)$$

と書ける。 $\sigma(c)=c$, $\mathbf{Z}_{it}=[y_{i,t-1}, \mathbf{X}_{it}]$ とすると上式は、式(9)のように通常分位点回帰モデルに書き換えることができる。

$$Q_Y(\tau|\mathbf{Z}_{it}) = \{\alpha_i + \delta_i F_\epsilon^{-1}(\tau)\} + \mathbf{Z}_{it}(\beta + \gamma)F_\epsilon^{-1}(\tau) \\ = \alpha'_{i,\tau} + \mathbf{Z}_{it}\beta'_\tau \quad (9)$$

積率条件から求めるべきパラメータは $\alpha_i, \delta_i, \beta, \gamma, F_\epsilon^{-1}(\tau)$ である。これらのうち、 $\alpha_i, \delta_i, \beta, \gamma$ は通常線形固定効果モデル同様 demean 変換によって推定値を得ることができる。 $F_\epsilon^{-1}(\tau)$ の推定値については、それらの推定値の結果から、分位点の計算だけで求めることができる。各パラメータの推定値から $\tau = 0.005, 0.015, \dots, 0.995$ の場合の条件付き分位点関数の推定値を式(10)によって計算する。

$$Q_Y(\tau|\widehat{\mathbf{Z}}_{it}) = \{\hat{\alpha}_i + \hat{\delta}_i F_\epsilon^{-1}(\tau)\} + \mathbf{Z}_{it}(\hat{\beta} + \hat{\gamma})F_\epsilon^{-1}(\tau) \quad (10)$$

この方法によって、大規模データであっても条件付き分位点関数を迅速に得ることができ、 $\sigma(\delta_i + \mathbf{X}_{it}\gamma) > 0$ である限り、交差の問題を回避することができ

る¹⁶⁾。

2. 部分線形変量効果モデル

代替モデルの第2の選択肢は、部分線形モデルに集団レベルの変量効果を導入するものである。変量効果によって観測されない異質性を十分に統制できるとは言い難いが、Bスプライン関数による過去の成績スコアの条件付けというSGPモデルのアイディアは保持することができる。分位点回帰モデルは、誤差項が非対称ラプラス分布に従う回帰モデルとして表現することができ、Kozumi&Kobayashi (2011)の混合分布に基づく方法を用いれば、ギブズサンプリングのみで事後分布からパラメータのサンプリングが可能である。またこの方法を用いれば、Normal-Gammaタイプの回帰モデルに帰着するため、スプライン関数を用いたセミパラメトリックモデル(部分線形モデル, 単一インデックスモデル), 変量効果モデルと組み合わせることも容易である。部分線形変量効果モデルは、式(11)として書くことができる。

$$\begin{aligned}
 y_{it} &= f_{\tau}(y_{i,t-1}) + \mathbf{X}_{it}\boldsymbol{\beta}_{\tau} + v_{i,\tau} + \epsilon_{it} \\
 \epsilon_{it} &\sim ALD(0, \sigma_{\epsilon}, \tau) \\
 v_{i,\tau} &\sim \mathcal{N}(\bar{\mathbf{X}}_i\boldsymbol{\alpha}_{\tau}, \sigma_{v_{\tau}})
 \end{aligned}
 \tag{11}$$

関数 $f(y_{i,t-1})$ の部分は、Bスプライン基底関数を用いることで線形モデルと同様の推定方法に帰着させることができる。マルコフ連鎖モンテカルロ法による各パラメータのサンプリングについては、Kozumi&Kobayashi (2011), Ruppert *et al.* (2003)の方法による。

条件付き分位点関数の交差の問題については、Rodrigues&Fan (2016)の方法によって解決することができる。Rodrigues&Fan (2016)は、非対称ラプラス分布に従う誤差項の回帰モデルについてマルコフ連鎖モンテカルロ法で発生させたサンプルに対して、事後的に変換を施すことで、交差しない分位点関数を形成することができる。この方法は、最初の非対称ラプラス分布の回帰モデルについて特定のモデリングの限定がない点、新たなサンプリングを行う必要がない点で有用な方法と言える。

C. シミュレーションデータによる比較

上記の分位点回帰モデルの拡張に関する議論をふまえて、シミュレーションデータに基づいて改めてSGPモデルとVAMの比較を行う。

シミュレーションでは、学校数=50, 各学校教員数=20, 各教員担任児童数=30, 計30,000人の児童のデータを2期分発生させた。各児童の現期のテストスコアは、前期のスコアと共変量 x , 観察されない要因に依存する。観察されない要因は、学校レベルの ν と児童レベルの ϵ に分解される。シミュレーションにあたっては、以下の3点において異なる計16の状況を設定する。

- A) 過去のスコアの寄与部分 $f(y_{ijs,t-1})$
 - 1) 線形のケース $f(y_{ijs,t-1})=0.9 y_{ijs,t-1}$
 - 2) 非線形のケース $f(y_{ijs,t-1})=20+60\Phi(1.5y_{ijs,t-1}^*)$
ただし、 y^* は、 y を標準化したものである。
- B) ν と児童レベル共変量 x の学校平均との間の相関
 - 1) 相関がないケース $\rho=0$
 - 2) 相関があるケース $\rho=0.4$
- C) ϵ の分布形
 - 式(12)の確率密度関数をもつ歪正規分布を想定し、歪度を司るパラメータ α について以下の4ケースを考える¹⁷⁾。
 - 1) $\alpha=0$
 - 2) $\alpha=1$
 - 3) $\alpha=5$,
 - 4) $\alpha=10$

$$f(\epsilon|\mu, \sigma, \alpha) = \frac{2}{\sigma} \phi\left(\frac{\epsilon - \mu}{\sigma}\right) \Phi\left[2\left(\frac{\epsilon - \mu}{\sigma}\right)\right] \tag{12}$$

また、SGPについては、以下の4点から、計8のモデルを推定する。

- a) 児童個人レベル共変量
 - 1) 共変量なし
 - 2) 共変量あり
- b) 学校効果
 - 1) 学校効果なし
 - 2) 学校効果あり (固定効果モデル)
 - 3) 学校効果あり (変量効果モデル)¹⁸⁾
- C) $f(y_{ijs,t-1})$ の関数形
 - 1) 線形
 - 2) 非線形 (Bスプライン)
- d) 個々の児童のSGPの教員レベルでの集計方法
 - 1) 平均値による集計
 - 2) 中央値による集計

表 1 各モデルの教員効果推定値と真値との相関

モデル	SGPモデル		固定効果SGPモデル		変量効果SGPモデル		固定効果VAM	2段階VAM	
	なし	あり	あり	なし	あり	なし			
共変量	なし	あり	あり	なし	あり	あり	—		
	なし	なし	なし	なし	あり	あり	—		
学校効果	なし	なし	なし	なし	あり	あり	—		
	なし	なし	なし	なし	あり	あり	—		
関数形	非線形	非線形	非線形	線形	非線形	非線形	—		
	(スプライン)	(スプライン)	(スプライン)	(スプライン)	(スプライン)	(スプライン)	—		
εの分布	内生性	平均値	中央値	平均値	中央値	平均値	中央値	中央値	
	ρ=0	0.299	0.318	0.867	0.808	0.901	0.816	0.808	0.840
正規分布 (α=0)	線形	Case 1	0.353	0.358	0.738	0.736	0.866	0.700	0.687
	ρ=0.4	Case 2	0.304	0.296	0.776	0.734	0.769	0.733	0.701
非線形	ρ=0	Case 3	0.280	0.275	0.775	0.721	0.798	0.547	0.710
	ρ=0.4	Case 4	0.445	0.449	0.801	0.765	0.839	0.765	0.806
線形	ρ=0	Case 5	0.416	0.405	0.842	0.788	0.836	0.841	0.859
	ρ=0.4	Case 6	0.393	0.393	0.785	0.755	0.789	0.782	0.716
非線形	ρ=0	Case 7	0.442	0.438	0.799	0.745	0.828	0.747	0.790
	ρ=0.4	Case 8	0.298	0.299	0.788	0.739	0.830	0.669	0.815
線形	ρ=0	Case 9	0.440	0.450	0.822	0.788	0.856	0.805	0.835
	ρ=0.4	Case 10	0.455	0.451	0.850	0.787	0.786	0.762	0.839
非線形	ρ=0	Case 11	0.381	0.380	0.867	0.827	0.841	0.863	0.845
	ρ=0.4	Case 12	0.286	0.273	0.792	0.738	0.843	0.756	0.841
線形	ρ=0	Case 13	0.223	0.242	0.717	0.680	0.834	0.726	0.812
	ρ=0.4	Case 14	0.340	0.340	0.839	0.784	0.832	0.827	0.824
非線形	ρ=0	Case 15	0.321	0.323	0.750	0.737	0.826	0.753	0.726
	ρ=0.4	Case 16							

学校数=50, 教員数1,000, 生徒数30,000としたシミュレーションデータ。
相関係数はSpearmanの順位相関係数

通常のSGPモデルは、a)については共変量なし、b)については学校効果なし、c)についてはスプライン関数を用いた非線形モデル、d)については中央値による集計、の組み合わせから成る。

表1は、シミュレーションデータにおける教員効果の真値と、各モデルにおけるその推定値との相関(Spearmanの順位相関係数)を示したものである。表に示された結果より以下の点が指摘できる。

まず第1に、共変量及び学校効果のない通常のSGPモデルでは、教員効果の推定値と真値との相関は低い。もちろん、この点は共変量と学校効果がスコアの分散をどれだけ規定するかというシミュレーションの設定に依存するが、どのケースにおいても他のモデルと比較しても相関は明らかに小さく、児童個人レベルの観測可能な共変量、学校レベルの要因を統制しないことによって、児童個人レベルの成長指標およびその集計に基づく教員効果は大きなバイアスを生む。既存研究においても指摘されていることではあるが、改めてプロトタイプのSGPモデルの用途は非常に限られることを強調しておきたい。

第2に、それ以外のSGPモデルの中では、固定効果SGPモデルの教員効果の推定値と真値との相関が高く、特に学校効果のないモデルとの比較においてはその差は大きい。また、固定効果SGPモデルと相関変量効果SGPモデルとでは、全てのケースで前者の固定効果モデルの方がより真値との相関が高い。この結果も非線形関数部分 $f(y_{ijt-1})$ の設定に依存するが、現期のスコアと過去のスコアとの相関が高いという経験的事実に照らせば、 $f(y_{ijt-1})$ が単調増加関数の形状から大きく乖離することを想定することは難しく、線形モデルであっても一定の妥当性があるのかもしれない。そして、固定効果SGPモデルは固定効果VAMと比較しても真値との相関において同等かそれ以上となっている。

これらから、本稿で提示した拡張SGPモデルは、プロトタイプのSGPモデルを大きく改善するものであり、教員の制御の及ばない要因の影響によるバイアスという難点に対処しうる。少なくともこの点において既存のVAMに劣後することはないと言える。

4. 考察

本稿では、近年のアメリカの教員評価・学校評価の評価手法として注目されてきたSGPモデルの方法的検討を行った。過去の成績に条件づけた相対的な成績

の情報を与える「成長指標」に基づく教員・学校評価モデルとしてはVAMが研究者間での長らく検討の対象となってきたため、SGPに関する検討は多くはなく、VAMに比較した評価についてもさほど高くないように思われる。

しかし、こうしたSGPモデルに対する低い評価はSGPモデルの内在的限界というよりも、その技術的中核をなす分位点回帰モデルに関する基礎的研究自体が少ないことに由来する。試論ながら本稿で示した拡張的モデルでは、これまで難点として指摘されてきた点の一部については解消可能であるように思われる。今後検討を要する事項は、セミパラメトリック回帰関数、非観測要因の統制、非交差制約を同時に満たすモデルの開発であり、パネル分位点回帰モデルに関する今後の研究の発展に依存するところが大きい。一方で、SGPモデルが依拠する分位点回帰モデルには、成績データの分布における外れ値・歪み、尺度変換に対する頑健性があり、こうした点の実用的利点は過小評価すべきではない¹⁹⁾。

本論では技術的な側面に紙幅を割いて論じたが、最後にVAMおよびSGPモデルに関する議論の教員評価・学校評価を含むアカウンタビリティ政策、あるいは教育政策の研究—社会関係に対する若干の含意を述べておきたい。

アメリカの教員評価・学校評価の評価手法は、教育生産関数研究で用いられてきたパネルデータに関する回帰モデルに由来し、その技術は教育経済学の実証研究や計量経済学の理論研究に依拠している。それらの企図として、教員、学校、あるいは自治体の努力・経営能力を他の要因から分離して評価すること、あるいは、VAMやSGPモデルの直観的なコンセプト自体については、統計学に関する知識の有無にかかわらず研究者と実務家の間に広く共有された理解が成立していると言える。

しかし、評価手法の統計的性質の評価に関する議論は時に専門的であり、そうした専門家の努力が評価の公平性を増進させるためのものであったとしても、評価の当事者や応用研究者の理解を大きく離れることになり、評価制度の正当性を揺るがす可能性も生じうる。一方で、1時点でのスコア、現期—過去の間の増分、過去の成績に条件づけた今期のランキングなどの簡便な記述的指標の学級・学校単位の集計量は理解が容易であるだけでなく、当事者の努力の動機づけに寄与しやすいのも事実であろう。すなわち、当事者から遠く離れた分析担当者によるデータ収集・分析を経て

初めて各当事者の寄与が明らかになるのではなく、当事者自身に具体的かつ即座に把握可能で、明確に意識されることによって——場合によっては、そうした簡便な記述的指標が公平性（因果推論としての妥当性）の面で難点があったとしても——結果的に期待された評価の順機能が作動しうる。こうした透明性と公平性のトレードオフの中で、評価制度に対する当事者間の合意がどのように形成されうるかという点については、純粋な方法的議論とは別に十分な考察が加えられるべきであろう。

注

- 1) 「成長」という語は、アメリカのテストベースの教員評価・学校評価の文献で頻出するが、明確な定義があるとは言い難い。本稿では、過去の成績に条件づけた現在の成績の位置という意味で用いる。
- 2) ただし、VAMにおいて関心の対象となるのは、通常のパネル固定効果・変量効果モデルでの個体効果であり、諸共変量における回帰係数ではない。
- 3) 成績推移表は、児童・生徒の成績をいくつかの順序カテゴリーに分類し、過去の時点の各成績カテゴリー該当者数と現期の各成績カテゴリー該当者数をクロスした表である。成績推移表には Progress Table, Value Tableなどの別名がある (Hewitt *et al.* 2016)。
- 4) SGPモデルは、「コラド成長モデル」と呼ばれることもある。
- 5) ただし、Lockwood&Castellano (2017) は、分位点回帰モデルではなく、共変量を伴う多次元項目反応理論において、成長指標および教員効果を推定する方法を提示している。この方法を適用するには、テストにおける項目別の解答データが必要である。
- 6) 条件付き分位点関数の記法については、分位点回帰モデルの慣例的な用法に従う。条件付き分位点を $Q_{\tau}(y|X)$ と記している文献もある。
- 7) Betebenner (2009, 2011) ではノット数を 4 とする 3 次スプライン関数を用いられている。4 個のノットは、 $y_{90,1}$ の 20, 40, 60, 80パーセンタイル点に置かれている。
- 8) 成績スコアの垂直的尺度化が必要となる成長指標について、単純な例としては、 $\Delta_{it} = y_{it} - y_{it-1}$ という差分が挙げられる。
- 9) この論文では、児童・生徒の成長指標に関して学級・学校単位の集計を行っていない。
- 10) 回帰関数が条件付き中央値となる中央値回帰は分位点回帰の特殊ケースであり、一般には回帰関数が条件付き期待値となる通常の回帰分析よりも外れ値に対して頑健であると言える。ただし、SGPでは、 $\tau = 0.005, 0.995$ といった極端な分位点での条件付き分位点も推定すること、Bスプラインによる非線形モデルとなっていることは、この頑健性に影響を与える。
- 11) 他にSGPとVAMについて比較した研究としては、Pivovarov&Amrein-Beardsley (2018), Sloat *et al.* (2018) など。
- 12) この 2 段階VAMは理論的に正当化されているとは言えないが、次の 2 点において大きな利点がある。まず、第 1 は、計算上の負荷がないことである (Koedel *et al.* 2015)。通常の固定効果モデル

で複数のレベルの固定効果がある場合には、計算負荷が非常に大きくなるが、2 段階法であればそれを回避することができる。第 2 は、教員（学校効果）を求めるときに、教員（学校）レベル変数をモデルに含めることができる点である。通常の固定効果モデルの場合、全てダミー変数に影響が吸収される。

- 13) 論文では、指導環境に対する中立性は「比例性」(proportionality) という単語で表現されている。「比例性」に関して、無償給食利用率と 3 つの学校レベル成果指標との相関係数は、SGPでは-0.37、1 段階VAMでは-0.25、2 段階VAMでは-0.03と報告されている。2 段階VAMにおいて、1 段階VAMと 2 段階VAMの相違は、後者が非時変的な学校レベル変数を明示的に統制している点である。
- 14) 1 段階VAMと 2 段階VAMとの比較に関しては、Ehlert *et al.* (2016) とは対照的に、前者の方が後者に勝るという結果を得ている。
- 15) 変量効果モデルでも時变的説明変数の級内平均を説明変数に加えれば、固定効果モデルと同じ結果を得ることができる。
- 16) $\sigma(\cdot) > 0$ を保証するためには、 $\sigma(\cdot) = \exp(\cdot)$ とするモデルが考えられる。この場合、demean変換とは異なるが、個体効果を除去することは可能である。
- 17) パラメータ α が 0 の時、正規分布となる。 α が 0 から離れるほど歪みのある分布となり、 α が負の時には左側に裾が長い分布に、正の値の時には右側に裾が長い分布となる。
- 18) 変量効果モデルとして、Abrevata&Dahl (2008) の相関変量効果モデルを用いた。
- 19) 被説明変数に外れ値や歪みを持つデータに頑健な回帰モデルとしては、誤差項に t 分布や歪正規分布を仮定したモデルが考えられる。

文献

- Abrevata, Jason and Christian M. Dahl 2008. "The Effects of Birth Inputs on Birthweight: Evidence from Quantile Estimation on Panel Data," *Journal of Business and Economic Statistics*, 26(4): 379-397.
- Arellano, Manuel and Stéphane Bonhomme 2016. "Nonlinear Panel Data Estimation via Quantile Regressions," *Econometric Journal*, 19: C61-C94.
- Ballou, Dale and Matthew G. Springer 2015. "Using Student Test Scores to Measure Teacher Performance: Some Problems in the Design and Implementation of Evaluation Systems," *Educational Researcher*, 44 (2): 77-86.
- Betebenner Damian 2009. "Norm- and Criterion-Referenced Student Growth," *Educational Measurement: Issues and Practice*, 28(4): 42-51.
- 2011. "A Technical Overview of the Student Growth Percentile Methodology: Student Growth Percentiles and Percentile Growth Projections/Trajectories"
- Bryk, Anthony S. and Herbert I. Weisberg 1976. "Value-Added Analysis: A Dynamic Approach to the Estimation of Treatment Effects," *Journal of Educational Statistics*, 1(2):127-155.
- Castellano, Katherine E. and Andrew D. Ho 2013. "Contrasting OLS and Quantile Regression Approaches to Student "Growth" Percentiles," *Journal of Educational and Behavioral Statistics*, 38(2): 190-215.
- and —— 2015. "Practical Differences Among Aggregate-Level Conditional Status Metrics: From Median Student Percentiles to Value-Add-

- ed Models," *Journal of Educational and Behavioral Statistics*, 40 (1): 35-68.
- Ehlert, Mark and Cory Koedel and Eric Parsons and Michael Podgursky 2016. "Selecting Growth Measures for Use in School Evaluation Systems: Should Proportionality Matter?," *Educational Policy*, 30(3): 465-500.
- Goldhaber, Dan and Joe Walch and Brian Gabele 2014. "Does the Model Matter? Exploring the Relationship Between Different Student Achievement-Based Teacher Assessments," *Statistics and Public Policy*, 1(1): 28-39.
- Guarino, Cassandra M. and Mark D. Reckase and Brian Stacy and Jeffrey M. Wooldridge 2015. "A Comparison of Student Growth Percentile and Value-Added Models of Teacher Performance," *Statistics and Public Policy*, 2(1): 66-76.
- Hanushek, Eric A. 1972. *Education and Race*, Heath, Lexington.
- He, Xuming 1997. "Quantile Curves without Crossing," *The American Statistician*, 51(2): 186-192.
- Hewitt, Kimberly Kappler and Audrey Amrein-Beardsley eds, 2016. *Student Growth Measures in Policy and Practice: Intended and Unintended Consequences of High-Stakes Teacher Evaluations*, Palgrave Macmillan.
- Koedel, Cory, Kata Mihaly and Jonah E. Rockoff 2015. "Value-Added Modeling: A Review," *Economics of Education Review*, 47: 180-195.
- Kozumi, Hideo and Genya Kobayashi 2011. "Gibbs Sampling Methods for Bayesian Quantile Regression," *Journal of Statistical Computation and Simulation*, 81(11): 1565-1578.
- Kurtz, Michael D. 2018. "Value-Added and Student Growth Percentile Models: What Drives Differences in Estimated Classroom Effects?," *Statistics and Public Policy*, 5(1): 1-8.
- Lockwood, J. R. and Katherine E. Castellano 2017. "Estimating True Student Growth Percentile Distributions Using Latent Regression Multidimensional IRT Models," *Educational Psychological Measurement*, 77(6): 917-944.
- Machado, José A. F. and J. M. C. Santos Silva 2019. "Quantiles via Moments," *Journal of Econometrics*, 213: 145-173.
- Muller, Jerry Z. 2018. *The Tyranny of Metrics*, Princeton University Press. (=2019, 松本裕訳『図りすぎ：なぜパフォーマンス評価は失敗するのか』みすず書房.)
- Murnane, R. J. 1975. *The Impact of School Resources on the Learning Inner City Children*, Ballinger, Cambridge.
- Pivovarov, Margarita and Audrey Amrein-Beardsley 2018. "Median Growth Percentiles (MGPs): Assessment of Intertemporal Stability and Correlations with Observational Scores," *Educational Assessment*, 23 (2): 139-155.
- Rodrigues, T. and Y. Fan 2017. "Regression Adjustment for Noncrossing Bayesian Quantile Regression," *Journal of Computational and Graphical Statistics*, 26(2): 275-284.
- Ruppert, David, M. P. Wand and R. J. Carroll 2003. *Semiparametric Regression*, Cambridge University Press.
- Sloat, Edward and Audrey Amrein-Beardsley and Jessica Holloway 2018. "Different Teacher-Level Effectiveness Estimates, Different Results: Inter-Model Concordance Across Six Generalized Value-Added Models (VAMs)", *Educational Assessment, Evaluation and Accountability*, 30(4): 367-397.
- Walsh, Elias and Eric Isenberg 2015. "How Does Value Added Compare to Student Growth Percentiles?," *Statistics and Public Policy*, 2(1): 53-65.
- Wright, Daniel B. 2018. "Estimating School Effectiveness with Student Growth Percentile and Gain Score Models," *Journal of Applied Statistics*, 45(14): 2536-2547.
- 〈付記〉
本研究はJSPS科研費JP19K02422, JP17H02664の助成を受けたものです。