

Title	博士課程進学に関する統計的因果探索の非線形効果の可視化
Author(s)	高山, 正行; 小松, 尚登; ファム, テトン; 前田, 高志ニコラス; 三内, 顕義; 小柴, 等; 清水, 昌平
Citation	年次学術大会講演要旨集, 40: 593-598
Issue Date	2025-11-08
Type	Conference Paper
Text version	publisher
URL	https://hdl.handle.net/10119/20148
Rights	本著作物は研究・イノベーション学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Research Policy and Innovation Management.
Description	一般講演要旨

博士課程進学に関する統計的因果探索の非線形効果の可視化

○ 高山 正行 (NISTEP/滋賀大学), 小松 尚登 (滋賀大学), ファム テ トン (滋賀大学/理化学研究所),
前田 高志 ニコラス (NISTEP/学習院大学/滋賀大学/東京電機大学/理化学研究所),
三内 顕義 (NISTEP/京都大学/滋賀大学/東京大学/理化学研究所/国立情報学研究所),
小柴 等 (NISTEP/滋賀大学), 清水 昌平 (NISTEP/大阪大学/滋賀大学/京都大学/理化学研究所)

1 はじめに

博士課程学生を含む若手研究者の育成・支援は、我が国の研究力強化の観点からも政策的に重要な要素であり、様々な調査・研究が行われてきた。著者らもこれまで、博士課程進学に関する因果探索を試み、議論を重ねてきた [高山 22, 高山 23, 高山 24]。近年では、大学の類型に応じて現れる因果効果の特徴の差異に関する解釈が進み、未観測共通要因の取り扱いも議論されている。さらに、信頼される因果モデルの構築や、政策的意思決定支援を念頭ににしたシミュレーションへの応用可能性・課題の整理もなされてきた [高山 24]。

ところで、これまでの一連の研究 [高山 22, 高山 23, 高山 24] では線形性を仮定しており、非線形性の考慮はなされていなかった。しかし、シミュレーション精度の向上を検討する上では、非線形性の考慮も重要な議論となる。そこで本研究では、これまでの議論を踏まえつつ、未観測共通要因の取り扱いを改善し、非線形な因果の影響についても推定・予測可能な計的因果探索アルゴリズムを応用した、より高度な因果探索・分析を試みる。加えて、推定された非線形の総合因果効果の可視化を通じ、因果メカニズムに関する議論の一層の深掘りを試みる。

2 研究手法

2.1 使用するデータ

本研究では、表 1 に示す通り、文献 [高山 24] で用いられた、国立大学における博士課程進学者数に関連する公開情報ベースのデータセットを、引き続き公開情報の範囲で改良・拡張した。

文献 [高山 24] では、統計的因果探索アルゴリズム FCI (Fast Causal Inference) アルゴリズム [Spirtes01] を適用した結果、「修士課程修了者数」と「博士課程修了者数」の間に、一定程度の未観測共通要因が存在する可能性を確認している。要因の候補としては、修士・博士の学位取得に必要な研究成果や、その創出に寄与する要因

に関する変数が想定されている [高山 24]。本研究では、その補完のため以下の通り、“Papers”、“KAKEN”の2種類の変数を追加した。その他の部分については、文献 [高山 24] と同様、大学改革支援・学位授与機構 (NIAD) が公開する大学基本情報^{*1}を中心にデータを構築した。

■Papers の追加 研究成果の指標として、オープンな研究成果書誌データベース “OpenAlex” [OpenAlex23] から、各大学に所属する著者を含む、各年の研究成果数を取得し、変数として追加した^{*2}。

■KAKEN の追加 研究成果の創出に寄与することが期待される指標には、既に変数として採用されている運営費交付金も含め、多様な指標が考えられる。ここでは、大学単位での研究活動を幅広く反映する指標として、日本学術振興会 (JSPS) による科研費の配分額を採用した。各年度の大学別配分額は、JSPS のホームページで公開されている情報から収集した。

以上を考慮の上、平成 24 年度から令和 4 年度までの各年度について、全 86 国立大学についてのデータセットを構築し、1 年度・1 大学あたり 1 点、データ点数としては合計 $11 \times 86 = 946$ 点となった。

2.2 因果探索アルゴリズム

本研究は以下の通り、文献 [高山 24] からのデータセットの拡充の効果検証を行ったうえで、非線形性を考慮した統計的因果探索アルゴリズム RESIT (Regression with Subsequent Independence Test) [Peters14] による因果探索・非線形効果の可視化を試みる。ただしいずれにおいても、2.3 節で示す分析条件に従う。

■FCI による拡張データセットの検証 文献 [高山 24] と同様、FCI と bootstrap 法を組み合わせ、データセットからの無作為復元抽出と因果探索を 1,000 回繰り返

^{*1} <https://portal.niad.ac.jp/ptrt/table.html>

^{*2} ただし、これは Institution について国立大学の名前のみを指定し、publication type は全てを含め、2012 年から 2022 年までの Stats から年単位の値を取り出したものである。そのため、年度単位でカウントするその他の変数とは基準が異なる点、そして有効数字は上から 4 桁目までに限られる点には、注意を要する。

表 1: 本研究で構築したデータセットの変数と出典元データ

変数名	内容	出典	変数構成の考え方
M_graduate	修士課程等修了者数	NIADのHPに掲載されている大学基本情報を基に加工	博士課程進学者数に関する変数
D_entrance	博士課程進学者数		
Ph.D_graduate	博士課程修了者数		博士課程修了者の修了直後のキャリアに関する変数
PosDoc	博士課程修了直後のポスト就職者数		
AcaPos	博士課程修了直後の大学教員就職者数	各国立大学法人の財務諸表等をもとに加工	大学の研究環境に係る変数
Basic_income	運営費交付金収益額		
Students_per_teacher	教員一人当たり学生数	NIADのHPに掲載されている国立大学の財務指標等を基に加工	博士課程進学者の経済的支援に係る変数
DC1	DC1採択者数	JSPSから過去に公開されたDC1採択者一覧のデータを基に加工	
KAKEN	科研費配分額	JSPSから過去に公開された機関別の配分額を収集	修士課程修了者数・博士課程修了者数の間に存在する未観測共通要因として考えられる候補
Papers	論文等の発表数	OpenAlexで当該機関所属の著者が一人でもいる文献をカウント	

本研究で新たに追加した変数

し、修士課程修了者数と博士課程修了者数の間に、未観測共通要因が示唆される辺が現れる bootstrap 確率を計測する。データセット拡張前（KAKEN, Papers なし）と拡張後（KAKEN, Papers あり）の両方で実験を行い、その bootstrap 確率に変化を検証する。また、文献 [高山 24] に準じた結果を得て比較するため、積の構造的因果モデルを仮定し、データセットの値は全て自然対数をとって分析する。なお、アルゴリズムは python package “causal-learn” [Zheng24] に公開されているものを使用した。

■RESIT による因果探索と非線形効果の可視化 RESIT は、以下の ANM (Additive Noise Model) を関数形として仮定した因果探索アルゴリズムである。本研究ではその中でも、最もシンプルで実践的とされる、以下の CAM (Causal Additive Model) [Bühlmann14] を仮定することとした。

$$x_i = \sum_{j \neq i} f_{ij}(x_j) + e_i \quad (1)$$

このとき、誤差関数 e_i に分布の制約を仮定せずとも、有向非巡回グラフであり、ある一定の微分可能性要件を満たせば、因果グラフが識別可能であることが知られている [Peters14]。RESIT ではこの性質を利用し、非線形性の回帰分析と、変数の残差同士の独立性検定^{*3}を行って、最も残差同士の依存性が少なくなる因果的順序を求めてから、その順序に基づいた非線形回帰分析と枝刈りを行うアルゴリズムとなっている。これは、DirectLiNGAM [Shimizu11] と類似した構造であり、かついずれも “lingam” [Ikeuchi23] で公開されている。

^{*3} ただし、Hilbert-Schmidt Independency Criterion に基づく、2 変数間の非線形な相関も捉える検定法を採用する。

以上に鑑み、本研究では “lingam” [Ikeuchi23] のアルゴリズムを使用し、DirectLiNGAM と RESIT での結果を比較する^{*4}。また、“lingam” [Ikeuchi23] には、bootstrap 法に基づいて、非線形の総合因果効果の可視化と統計的信頼性の評価を行う機能が実装されているため、1,000 回の bootstrap sampling を行い、この機能を活用した分析を実施する。

なお、本パッケージの RESIT では、非線形回帰分析のアルゴリズムも自由に設定できるが、本研究では CAM の仮定に従い、線形一般化加法モデルに基づく “linearGAM” を採用した。

2.3 分析にあたっての条件設定の詳細

■経時変化を加味する範囲 遅延時間依存性は、本研究では 1 年度分まで考慮に入れることとした。つまり、0 年目である時刻 t (年度) とその 1 年後である $t+1$ (年度) の 2 時点を考慮して、時間的因果律も考慮した経時変化を見る。このため、実際の分析に当たっては、すべての変数の 1 年後の値について、変数名 + “_L1” と区別したうえで、データセットに加えて分析する。

■国立大学の重点支援 3 類型に応じた層別化 国立大学の性質に応じた層別化は、国立大学の第 3 期中期目標期間 (～2021 年度) まで、文部科学省により使われていた、運営費交付金の重点支援の 3 類型に基づいて実施し、特に以下の類型 I と類型 III について計算・比較を行う。

- 重点支援類型 I: 地域に貢献する取組を行い、専門分野の特性に配慮し、強み・特色のある分野で世界・

^{*4} この際 DirectLiNGAM では、文献 [高山 23, 高山 24] との連続性を重視し、データセットの値は全て自然対数を取り、積の構造的因果モデルを仮定して分析するが、RESIT は指数関数の挙動も捉えられるアルゴリズムのため、自然対数はとらない。

全国的な教育研究を推進する国立大学（全 55 大学）

- 重点支援類型 II: 専門分野の特性に配慮しつつ、強み・特色ある分野で地域というより世界全国的な教育研究を推進する国立大学（全 15 大学）
- 重点支援類型 III: 卓越した成果を創出している海外大学と伍して、卓越した教育研究、社会実装を全学的に推進する国立大学（全 16 大学）

■事前知識の設定 本研究で活用する、“causal-learn”の FCI, “lingam”の DirectLiNGAM と RESIT は、事前知識 (prior knowledge) が設定可能である。因果探索の制約条件を適切に設定し、領域知識の観点からはありえない向きの辺が現れるのを防止しつつ、自然な向きの辺が現れるように制御できる。本研究では、変数の増加に伴い、文献 [高山 23, 高山 24] より、解釈も煩雑になるため、効率化に向けて以下の通り、事前知識を設定した。

(A) 同じ年度において内数の関係にあるパスの設定 例えば、博士課程進学者数は、統計上、修士課程修了者数の内数なので、「D_entrance → M_graduate」は不自然なパスである。同様に、「Posdoc, AcaPos → Ph.D_graduate」も、統計上の内数の関係に鑑み、不自然なパスである。よってこれらのパスは禁止とした。

(B) 同じ年度での DC1 と他の変数の関係の設定 DC1 の採択者数は、その年度の半年近く前にほぼ固まっており、他の変数に比べ値の確定が明らかに早い。よって、同じ年度内での DC1 への他の変数からのパスは禁止とした。

(C) 運営費交付収益額と他の変数の関係の設定 文部科学省の運営費交付金の配分基準^{*5}では、前年度の運営費交付金配分額(≒Basic_income)をベースに、D_entrance, Ph.D_graduate, PosDoc, AcaPos, KAKEN, Papers ^{*6}に依じて、後年度の運営費交付金の配分額が決まることから、これらの変数からの Basic_income_L1 への影響は、枝刈りで消える場合を除き、必ず現れるものとした。

(D) 未来からの影響の禁止 未来から過去へ影響するパス(“L1”がついた変数から、ついていない変数へのパス)は全て禁止とした。

3 結果・考察

3.1 FCI における未観測共通要因の存在可能性

表 2 に、FCI と bootstrap 法を組み合わせた実験結果を示す。データセットに KAKEN と Papers を追加した結果、M_graduate と Ph.D_graduate の間に未観測共

表 2: FCI と bootstrap 法を組み合わせ、M_graduate と Ph.D_graduate の間に未観測共通要因が示唆される確率。なお、FCI で出力される辺は、以下のように定義される。

$A \leftrightarrow B$: A と B は互いに相手の原因にはならないが、未観測共通原因がある。

$A \circ - \circ B$: 『A が B の原因』, 『B が A の原因』, 『A と B の間に未観測共通原因が存在する』のうち少なくとも 1 つが成立。

	データセットに KAKEN, Papersなし	データセットに KAKEN, Papersあり
類型 I M_Graduate とPh.D_Graduateの間	○-○ bootstrap確率0.234	○-○ bootstrap確率0.109
類型 I M_Graduate_L1 とPh.D_Graduate_L1の間	↔ bootstrap確率0.143	↔ bootstrap確率0.104
類型 III M_Graduate とPh.D_Graduateの間	○-○ bootstrap確率0.111	完全に分離

通要因の存在が示唆される bootstrap 確率は、類型 I では減少し、類型 III では完全に消失した。このことから、FCI で統計的に示唆された未観測共通要因は、KAKEN や Papers と関係していた可能性が高いと考えられる。一方、類型 I では確率が完全には消失せず、いずれも 0.1 以上を示しており、他にも未観測共通要因が存在する可能性を示唆している。

また、紙面の都合上掲載は割愛するが、データセット全体に対する FCI での因果探索では、文献 [高山 24] とは異なり、M_graduate と Ph.D_graduate の間の未観測共通要因が示唆されるパスは、KAKEN と Papers の追加で消失した。この点からも、データセットの改善効果が確認できる。もっとも、両変数間の問題を完全に解消することを目指すとは、逆に分析精度が低下する懸念も残る。

以上を踏まえ、M_graduate と Ph.D_graduate の間の未観測共通要因に関する問題は、KAKEN と Papers の追加により十分に改善されたと判断し、RESIT による分析を進めることとした。

3.2 RESIT で見られた因果グラフの全体像と DirectLiNGAM の結果との比較

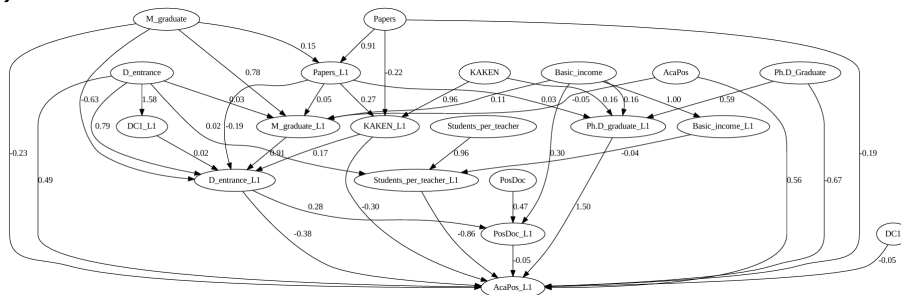
図 1 には、類型 III のデータセット全体に対する、(a) DirectLiNGAM で積の構造的因果モデルを仮定して計算した結果と、(b) RESIT で計算した結果を示す。ここで、(a) の辺の値は、線形構造方程式の因果係数(積の構造的因果モデルの場合は、厳密には指数)の推定値である。他方、(b) の辺の値は全て 1.00 となっており、これは何らか非線形関数でフィットされ、枝刈りを受けず、その変数を含む非線形関数が残存することを意味する。

全体的な相違点として、(b) の RESIT による因果グラフの方が、(a) の LiNGAM による因果グラフより辺の数が多く、複雑な構造になっている。これは、LiNGAM で

^{*5} www.mext.go.jp/a_menu/koutou/houjin/1417427.htm

^{*6} 厳密には、これらの変数の過去 3 年分の平均値が評価されるが、本研究では、1 年前の変数からの影響で近似できると仮定した。

(a) DirectLiNGAM (積の構造的因果モデルを仮定)



(b) RESIT

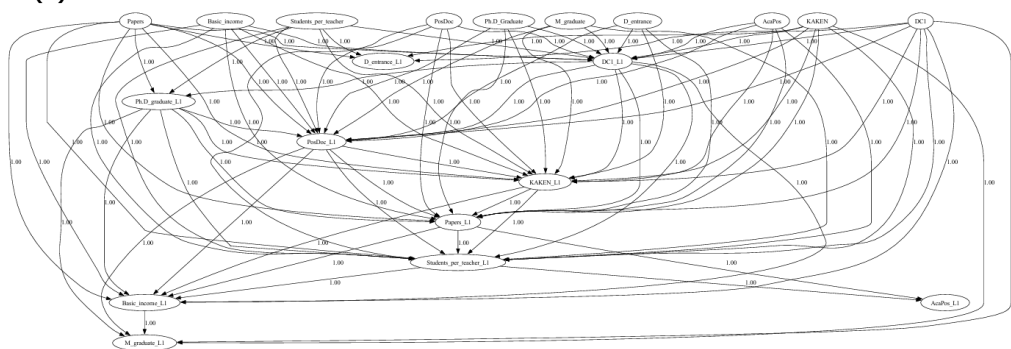


図 1: 重点支援類型Ⅲの大学に関し、データセット全体に対して因果探索した結果の比較。(a) DirectLiNGAM で、積の統計的因果モデルを仮定し、データセット全体に対して自然対数をとったうえで計算した結果。(b) RESIT で、データセット全体をそのまま使って計算した結果。

は線形のフィットで十分に表現しきれず、結果的に限られた関係のみが抽出されている一方、RESIT では非線形性が考慮されることで、フィッティングや独立性検定での感度が高まり、枝刈りの影響を受けにくくなったからと考えられる。

ただし、RESIT の結果は LiNGAM に比べて解釈が複雑であり、かつ LiNGAM の因果係数のように、効果の正負や大きさを直接示す情報を持たない。したがって、因果グラフのみでは定量的な議論が困難である。そのため、特定の変数に介入した場合に、他の変数がどう変化するか、総合因果効果の可視化やシミュレーションを通じて評価し、その統計的信頼性を議論することが重要である。

3.3 非線形因果効果の可視化

“lingam”[Ikeuchi23] に実装された機能に基づく非線形効果の可視化の結果のうち、代表的なものを図 2 に示す。具体的には、bootstrap 標本ごとに、RESIT で因果探索した結果に基づき、変数 x_j に介入を行って $x_j = a$ と固定した際の変数 x_i への総合因果効果 $E(x_i | do(x_j = a))$ ^{*7} を計算し、横軸を x_j 、縦軸を $E(x_i | do(x_j = a))$ とし

て、白丸でプロットした。この時、介入での固定値 a は、データセット全体の値域で 9 等分し、10 点とした。また、緑線は bootstrap 標本での $E(x_i | do(x_j = a))$ の平均値を繋いだ線であり、さらに信頼区間片側 5% の推移を、上側を赤線、下側を青線で表現した。これにより、信頼区間にも留意の上、 $E(x_i | do(x_j = a))$ の挙動を掴める。

本研究では、文献 [高山 23, 高山 24] において、類型Ⅲにおいて直接のパスが特に顕著にみられるとされた、 $DC1 \rightarrow D_entrance$, $D_entrance \rightarrow DC1.L1$ に加え、今回新たに変数を追加したことで観測される、 $KAKEN \rightarrow DC1.L1$ を、図 2 に代表的に取り上げている。

■ $DC1 \rightarrow D_entrance$ これは、 $DC1$ のような経済的支援によって博士課程進学者数が増えるという、自然に解釈しやすいパスである。文献 [高山 23, 高山 24] では、FCI や DirectLiNGAM において、類型Ⅲでは直接の正の影響が観測されやすい一方、類型Ⅰでは直接の影響が現れにくいとされていた。図 2(a) を見ると、類型Ⅰでは $D_entrance$ の期待値の、 $D_entrance$ の増加に伴う変化は、微増傾向であるものの、比較的すぐに頭打ちとなる。一方図 2(d) の通り、類型Ⅲでは $D_entrance$ の期待値は、 $DC1$ の増加に伴い概ね線形で増加している。よって、RESIT による図 2(a)・(d) では、文献 [高山 23, 高山 24] の傾向が、より高い感度で検出されており、解釈も整合

^{*7} データセット内の別の変数 x_k を介して伝搬する効果も含む。そのため、直接のパスの bootstrap 確率が小さくとも、 $E(x_i | do(x_j = a))$ の変化が小さくなるわけではない。

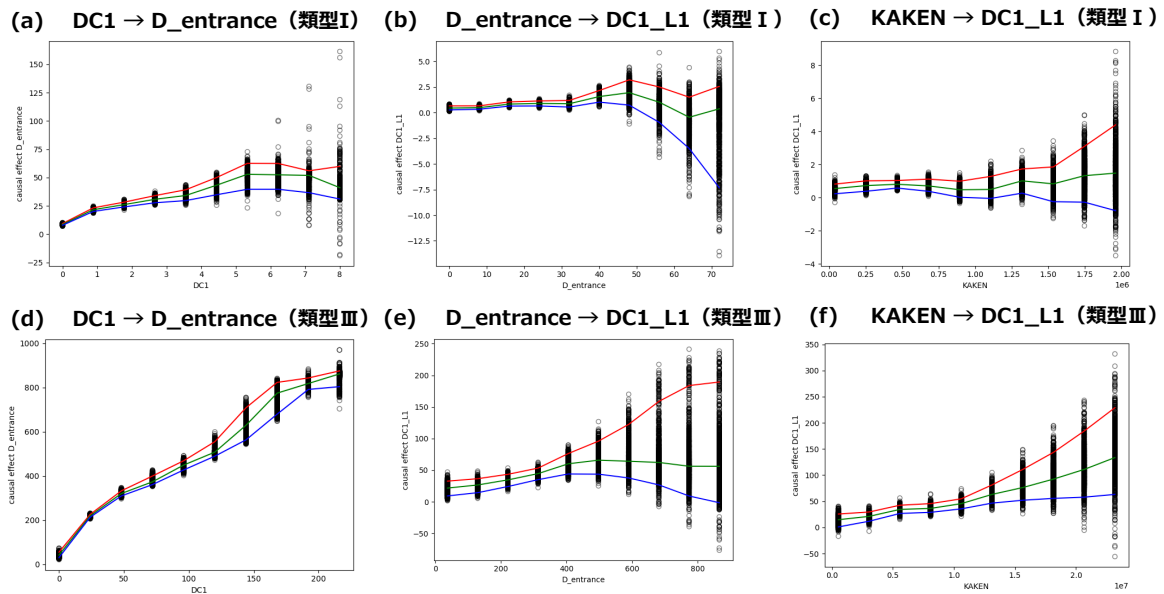


図 2: RESIT を bootstrap 法と組み合わせて因果探索・予測を繰り返し、代表的なパスについて非線形因果効果を可視化したもの。

する。また、いずれの場合も信頼区間は、DC1 の増加に伴い極端に大きく広がってはいない。

■D_entrance→DC1.L1 文献 [高山 23, 高山 24] では、FCI や DirectLiNGAM において、類型 III では直接の影響が正の影響として、しかも DC1 の 2 次以上の高次の影響として観測される一方、類型 I では直接の影響が現れにくいものとされる。このパスは、仮説ではあるが、類型 III のような研究大学では、直近で博士課程に進学した先輩学生が多いことで、1 学年下の修士学生の DC1 採択の機会増に繋がるというメカニズムで解釈されている。実際、図 2(b)・(e) を見ると、類型 I では DC1.L1 の期待値について、D_entrance の増加に伴う変化はほぼ見られないが、類型 III では DC1.L1 の期待値が、特に D_entrance が 100 以下の範囲で増加傾向であり、しかも僅かに下に凸にある。よって、RESIT による図 2(b)・(e) でも、文献 [高山 23, 高山 24] と整合する結果を得られた。

一方、類型 III における D_entrance の値が 100 以上の範囲での振る舞いは、信頼区間も不安定であり、平均を示す緑線も飽和傾向がみられる。この飽和現象は、FCI や DirectLiNGAM の範囲の分析では観測し得なかったものであり、RESIT による非線形効果を考慮した分析で初めて観測されたものである。上述の D_entrance→DC1.L1 のメカニズムの仮説に基づけば、博士課程に進学した先輩学生が多くなっていても、1 学年下の修士学生と実際つながる先輩学生の数自体は線形で増えていくわけではなく、また DC1 採択に向けて実際に得られる助言等の機会が増えたとしても、その有効性は頭打ちになって

いるものと考えられる。

■KAKEN→DC1.L1 図 2(c) の通り、類型 I では、KAKEN の増加に伴い、信頼区間の広がり観測されるものの、DC1.L1 の期待値はほぼ変化していない。一方、図 2(f) の通り、類型 III では、KAKEN の増加に伴い、DC1.L1 の増加傾向が見られる。特に類型 III のような研究大学においては、科研費の取得が上手くいっている部門では、そのノウハウが、学生の DC 1 採択に向けた申請書執筆における、研究計画・提案の質の向上にもつながり、結果的に多くの採択にも繋がっていることによるものと解釈される。

4 おわりに

4.1 本研究での成果のまとめ

本研究ではまず、国立大学における博士課程進学に関するデータセットを、先行研究 [高山 24] を拡充する形で再構築し、未観測共通要因の存在・影響をより適切に捉えられるよう改善した。これにより、研究力強化の文脈に即した議論が可能となった。

次に、非線形因果効果を考慮可能な統計的因果探索アルゴリズム RESIT を適用し、非線形な総合因果効果を可視化した。その結果、以下の知見を得た。

- DC1 による博士課程進学者数への影響が、より高い感度で分析できたこと
- 類型 III の研究大学において、博士課程進学者数の増加が翌年の DC1 採択者数増になるが一方、その効果に飽和現象が示唆されたこと

- ・ 類型Ⅲの研究大学において、科研費の配分額がDC1採択者数増に影響を与える可能性が観測・解釈されたこと

これらにより、非線形的な挙動を踏まえた、より精緻な分析と解釈が可能となった。

4.2 今後の政策面での貢献の可能性

本研究で示した通り、RESITの活用は、非線形因果効果について、因果メカニズムに関する解釈の深堀を可能にする。さらに、この非線形因果効果の可視化は、信頼区間の評価も行っており、リスクも含めた精緻な予測シミュレーションにも寄与すると期待される。この点で政策的意義も大きい。

加えて、本研究で得られた結果・解釈は、各大学の環境や状況に鑑みた、博士課程進学者数の増加や研究力強化に向けた、個別最適な支援の在り方の議論にも資する。また、この因果探索に基づいたシミュレーションにより、定量的・統計的な裏付けを伴う政策議論の展開も可能となり、科学技術イノベーション政策への貢献が期待される。

4.3 本研究の限界と今後の展望

一方で、本研究には以下の限界があり、今後はその克服を通じて応用の深化を図る必要がある。

■ **因果グラフの各辺の妥当性の網羅的検証** 本研究では、変数間の因果関係（辺）の妥当性について網羅的な検証は行っていない。信頼性の高い因果モデルを構築するには、向きの妥当性を含め、統計的・理論的に検証することが不可欠である。近年は、大規模言語モデルと統計的因果探索を組み合わせ、領域知識と統計的妥当性の両面を担保する方法 [Takayama25] が提案されており、こうした手法の活用も有望である。

■ **因果モデルを用いたシミュレーション** 非線形性を考慮した因果モデルを構築することで、より高い精度で、具体的な政策意思決定支援への展開が期待される。今後は、“analysis”から“synthesis”にも踏み込み、例えばマルチエージェントシミュレーション等も活用し、複数の政策シナリオを構築・比較する応用へ発展させていくことが課題である。

本研究の分析コード・結果の公開

本研究の分析のpythonコードや、RESIT等で得られた因果グラフ、全ての変数の組み合わせでの総合因果効果の可視化等の結果は、以下に公開している。

<https://github.com/mas-takayama/Causal-Discovery-in-Doctoral-Problems-in-Japan/tree/main/JSRPIM40th%202B03>

謝辞

本研究の一部は、JST、CREST、JPMJCR22D2の支援を受けて実施したものである。

参考文献

- [高山22] 高山正行, 小柴等, 前田高志ニコラス, 三内顕義, 清水昌平, 星野利彦: 博士課程進学率に関する因果モデルの構築: 統計的因果探索アルゴリズム“LiNGAM”による試行的分析. *Jxiv*, 2022. (preprint) <https://doi.org/10.51094/jxiv.1>
- [高山23] 高山正行, 小松尚登, ファム テトン, 前田高志ニコラス, 三内顕義, 小柴等, 清水昌平: 大学別の博士課程進学等に関するデータセットの構築と統計的因果探索. **研究イノベーション学会 第38回年次学術大会（予稿集）**, 公演番号 2D20, 2023.
- [高山24] 高山正行, 小松尚登, ファム テトン, 前田高志ニコラス, 三内顕義, 小柴等, 清水昌平: 博士課程進学者数に関する統計的因果探索と交絡因子の取り扱い. **研究イノベーション学会 第39回年次学術大会（予稿集）**, 公演番号 1D06, 2023.
- [OpenAlex23] Jason Priem, Heather Piwowar, and Richard Orr: OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv*, 2024. (preprint) <https://doi.org/10.48550/arXiv.2205.01833>
- [Shimizu06] Shohei Shimizu, Patrik O. Hoyer, Aapo Hyvärinen, and Antti Kerminen: A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7:2003-2030, 2006. <https://www.cs.helsinki.fi/group/neuroinf/lingam/JMLR06.pdf>
- [Shimizu11] S. Shimizu, T. Inazumi, Y. Sogawa, A. Hyvärinen, Y. Kawahara, T. Washio, P. O. Hoyer and K. Bollen.: DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model. *Journal of Machine Learning Research*, 12(Apr): 1225-1248, 2011. <https://dl.acm.org/doi/10.5555/1953048.2021040>
- [Ikeuchi23] Takashi Ikeuchi, Mayumi Ide, Yan Zeng, Takashi Nicholas Maeda, and Shohei Shimizu: Python package for causal discovery based on LiNGAM. *Journal of Machine Learning Research*, 24(14):1-8, 2023. <http://jmlr.org/papers/v24/21-0321.html>
- [Peters14] J. Peters, J. M. Mooij, D. Janzing, and B. Schölkopf: Causal discovery with continuous additive noise models. *Journal of Machine Learning Research*, 15(58):2009-2053, 2014. <http://jmlr.org/papers/v15/peters14a.html>
- [Bühlmann14] P. Bühlmann, J. Peters, and J. Ernest: CAM: Causal additive models, high-dimensional order search and penalized regression. *The Annals of statistics*, 42(6):2526-2556, 2014. <http://jmlr.org/papers/v15/peters14a.html>
- [Zheng24] Y. Zheng, B. Huang, W. Chen, J. Ramsey, M. Gong, R. Cai, S. Shimizu, P. Spirtes, K. Zhang: Causal-learn: Causal discovery in Python. *Journal of Machine Learning Research*, 25(60):1-8, 2024. <https://www.jmlr.org/papers/v25/23-0970.html>
- [Spirtes01] Peter Spirtes: An Anytime Algorithm for Causal Inference. *Proceedings of Machine Learning Research*, R3:278-285, 2001. <https://proceedings.mlr.press/r3/spirtes01a.html>
- [Takayama25] Masayuki Takayama and Tadahisa Okuda and Thong Pham and Tatsuyoshi Ikenoue and Shingo Fukuma and Shohei Shimizu and Akiyoshi Sannai: Integrating Large Language Models in Causal Discovery: A Statistical Causal Approach. *Transactions on Machine Learning Research*, 2025. <https://openreview.net/forum?id=Reh1S8rxfh>