JAIST Repository
https://dspace.jaist.ac.jp/

Title	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
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, 38: 874-879
Issue Date	2023-10-28
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/19271
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

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

２Ｄ２０

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

○ 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/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）,
小柴 等（NISTEP/滋賀大学）, 清水 昌平（NISTEP/滋賀大学/理化学研究所）

1　はじめに

科学技術・イノベーション (STI) 政策において若手研究者支援は重要なテーマであり，特に大学院博士課程の進学支援等に関しては近年，予算的にも重点的な措置が取られている。また，これらが EBPM の文脈とも相まって STI 政策研究の主要テーマの一つとしても様々に取り上げられている。例えば NISTEP は，博士人材追跡調査や修士課程（6 年制学科を含む）在籍者を起点とした追跡調査を一連のコホート調査として実施しており（例えば，[治部 21]），博士課程進学に必要とされる要因や修了後の進路に関する知見が蓄積されている。

ところで，「人材」について直接的に扱ってきた学問分野は経済学や経営学など社会学系であり，上述の各種博士課程進学関連の研究についても，そうしたディシプリンに基づいて行われるもの，あるいは，より実務的に政策的な領域知識に基づいて行われるものが多かった。これに対して近年では，データドリブンな分析，例えば情報理論や統計学の枠組に基づく統計的因果探索による分析も試みられている [高山 22a, 高山 22c]。中でも特に，構造的因果モデル (Structural Causal Model, SCM) に基づき，データ駆動的に尤もらしい有向非巡回グラフ (Directed Acyclic Graph, DAG) の出力を可能にする統計的因果探索アルゴリズムの一つである LiNGAM は，その後の SCM ベースでの議論に基づき構築された因果グラフでデジタルツイン的に input/output についてのシミュレーションも可能にし得るツールとして着目され，博士課程進学に関する政策研究領域でも応用が試みられてきている [高山 22a, 高山 22b]。これらの研究はいずれも国全体の状況を把握することを目的に，国の統計値レベルでのオープンデータを活用し，因果探索を試みたものであるが，「進学」というイベントが基本的には年の単位で発生・記録されるものであり，かつ，「国全体」の統計値を用いているため，要因と想定され得る変数の数に対し，利用できるデータ点数が厳しく制限されるという課題がある。そのため時系列的な影響が十分に考慮できていなかったり，大量の要因からボトムアップに可能性を探索するデータ駆動の利点を十分に生かし切れていないなど，分析の限界も指摘されてきている [高山 22b]。

この課題に対する最も単純な対応として，例えば大学別，分野別等といった形でデータセットを構築することで，データ点数を増やしていくというアプローチが考えられる。また，手法自体を改良するというアプローチも考えられる。

これらの背景を踏まえ，本研究では公開データから大学別のデータセットを再構築し，さらに将来的には理学・工学などの分野別かつ遅延時間も考慮した因果探索が可能となるよう，工夫を試みた。本稿ではその方法と計算結果についてまとめる。

2　研究手法

2.1　使用するデータ

本研究ではまず，既存の博士課程進学に関する因果探索の取組 [高山 22a] の考え方を参考に，（独）大学改革支援・学位授与機構 (NIAD) によって公開されている学校基本情報を中心としてデータを構築する。NIAD で公表しているこれらの情報は，平成 24 年度から令和 4 年度までの各年度について，国公立大学・短期大学から提供された基礎的な情報を掲載したものであり，文部科学省の学校基本調査への回答に基づくものとなっている。そのため，大学別の修士課程・博士課程の入学者数や進路別の修了者数等について，変数に組み込むことが可能となる。

NIAD の学校基本情報のデータを中心に，以下に示す通り，(1) 博士課程進学率に係る変数，(2) 博士課程修了者の修了直後のキャリアに係る変数，(3) 大学の研究環境に係る変数，(4) 博士課程進学直後の経済的支援，に係る変数の 4 つの観点から変数を合計 8 つ選択し，表 1 の通り，各出典元データから加工・突合してデータセットを構築した。

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

変数の種類	出典元データ
修士課程修了者数[*1]	NIAD 学校基本情報[*2] を基に加工
博士課程進学者数	
博士課程修了者数	
博士課程修了後の ポストドクター就職者数	
博士課程修了後の 大学教員就職者数	
運営費交付金	文部科学省 各目予算明細書[*3] を基に加工
教員一人当たり学生数	NIAD 国立大学法人の 財務指標等を基に加工
JSPS特別研究員DC1採択者数	JSPSのデータを基に加工

(1) **博士課程進学率に係る変数** 博士課程進学率は 0 から 1 の範囲に収まる変数であり，少なくとも直接線形性を仮定して構造方程式モデルで因果推論を行うことには馴染まない。そこで，進学率そのものを変数とするのではなく，その算出に用いられる，

- 修士課程修了者数
- 博士課程進学者数[*4]

を使用する。

(2) **博士課程修了者の修了直後のキャリアに係る変数** 進学率と同様に，まずは博士課程修了者数というキャリアを考える上での母数に着目し，さらに修了後に想定される代表的なキャリアの一つがアカデミアであることを考慮して，本研究では，

- 博士課程修了者数
- 博士課程修了直後のポストドクター等への就職者数[*5]
- 博士課程修了直後の大学教員への就職者数[*5]

を採用している。

(3) **大学の研究環境に係る変数** 大学の研究環境の指標の候補となる変数は様々考えられるが，例えば国から配分される基盤的経費について，国立大学と私立大学では運営費交付金が存在の有無が異なるなど，大学種に応じて前提が大きく異なる異なることが想像される。そこで，ここでは分析の対象を一旦，国立大学のみに絞ることとし，また公開データに基づいて，NIAD が公開する学校基本情報と同じ粒度で入手可能な情報として，

- 運営費交付金の配分額
- 教員一人当たり学生数

を採用する。なお，博士課程進学の課題に対して統計的因果探索を適用した先行研究等 [高山 22a, 高山 22b] では，研究環境の変数の一つとして，教員の研究時間割合を採用していたが，これは 5 年に 1 回程度実施されている「大学等におけるフルタイム換算に関する調査」[*6]に基づくものであり，そもそもこの調査が抽出で実施されていることもあって，大学単位で毎年のデータにアクセスすることはできない。そこで本研究では，毎年度大学ごとに公開されている情報を基に算出可能であり，かつ教員の忙しさを示しうる代替指標として，教員一人当たりの学生数を採用した。

(4) **博士課程進学直後の経済的支援に係る変数** 経済的支援の重要性は，科学技術・イノベーション基本計画においても，そして博士課程進学に関する様々な調査においても指摘されてきており，博士課程進学者数に対する原因変数の有力な候補と考えられる。一方，経済的支援は種類・規模・1 件当たりの支援額が様々であり，少なくとも毎年度大学ごとのデータとして把握することが難しいものも多い。そこで本研究では，経済的支援の代表例の一つである日本学術振興会 (JSPS) の特別研究員制度に着目し，特に，博士課程進学前に採択が定まる，

- JSPS の特別研究員 DC1 採択者数

のみを暫定的に採用する。

以上の考え方により，平成 24 年度から令和元年度までの 9 か年度にわたっての国立大学全 86 について，これらの 8 変数でデータセットを構築した。1 年度・1 大学あたり 1 点，データ点数としては合計 $9 \times 86 = 774$ 点となっている。先行研究 [高山 22a, 高山 22b] では，ほぼ同程度の変数に対してデータ点数が 10 点程度であったが，これらと比べてデータ点数が約 70 倍以上となった。

ただし留意点も存在する。現状では 86 の異なる国立大学について同質なもの・対等なものとしているが，実際には研究に力を入れている大学もあれば，教育が中心

[*1] 医学部・薬学部の 6 年制カリキュラムの修了も，ここに含めている。

[*2] https://portal.niad.ac.jp/ptrt/table.html

[*3] 例えば令和元年度については，https://www.mext.go.jp/component/b_menu/other/__icsFiles/afieldfile/2019/02/13/1413363_01_1.pdf を参照

[*4] ただし，ここでいう博士課程進学者数は，あくまで各大学の修士課程修了者のうち，進路として博士課程等へ進学予定とした者のことであり，進学した博士課程の大学が，修了した修士課程の大学と一致しないこともあることには留意する必要がある。

[*5] 博士課程進学者と同様，あくまで各大学の博士課程修了者のうち，これらの進路に進む予定とした者であるため，博士課程を修了した大学と就職先の大学が一致するとは限らない。

[*6] https://www.mext.go.jp/b_menu/toukei/chousa06/fulltime/1284874.htm

となっている大学もあるなど，本来はそれぞれ強みも環境も異なり，博士課程進学者の割合などにも差がある。したがって，大学の性質により因果関係も異なる可能性がある。つまり，現状，このデータセット全体で見ると国立大学全体を捉えていることになるものの，データ点一つ一つを見ると，単純な比較が成り立たないところもある。

2.2 因果探索のアルゴリズム

2.1 で述べたデータの性質と，1 章で述べた趣旨に鑑み，Longitudinal LiNGAM[Kadowaki13] のアルゴリズムを出発点として因果探索の方法を検討する。このアルゴリズムは，以下に示す式の通り，時系列データに対して，同時点の因果効果に加えて時間遅れを伴って表れる因果効果も含め，さらにその因果効果そのものの経時変化についても考慮した SCM による因果探索を行うものである。

$$x_i(t) = \sum_j \sum_{\tau=0}^{\tau_{max}} b_{i,j}(t, t-\tau) x_j(t-\tau) + e_i(t) \qquad (1)$$

時刻 t における i 番目の変数 $x_i^{(g)}$ を，時刻 t 以前の変数による線形構造方程式で表すものであり，$b_{i,j}(t, t-\tau)$ は $x_j(t-\tau)$ の変化が，時間 τ の遅れを伴って $x_i(t)$ に及ぼす影響を表す因果係数*7，また $e_i(t)$ は $x_i(t)$ の誤差変数となっている。

そして本稿では，この Longitudinal LiNGAM を拡張し，時点が変わっても同時点の因果的順序は共通であるという制約*8が入るようにした。これにより，いよいよ 2.1 で構築したデータセットの分析の準備が整う。

2.3 分析にあたっての条件設定

2.1 と 2.2 に基づいた分析を行う際の条件として，以下の通り設定する。

■扱う変数の定義と経時変化を加味する範囲　扱う変数は 2.1 に述べた全ての変数を採用する。これに伴う変数の表記法については表 2 にまとめた。また，遅延時間依存性については，今回は 2 年度分まで入れることとする。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)，そして 2 年後の $t+2$(年度) の 3 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。

表 2: 本研究での分析における各変数の定義*9

変数に関する説明	変数名
修士課程修了者数	$x_0(t)$
博士課程進学者数	$x_1(t)$
博士課程修了者数	$x_2(t)$
博士課程修了後のポストドクター就職者数	$x_3(t)$
博士課程修了後の大学教員就職者数	$x_4(t)$
運営費交付金	$x_5(t)$
教員一人当たり学生数	$x_6(t)$
DC1採択者数	$x_7(t)$

■積の構造的因果モデルの仮定　本研究のような事例では，例えば「修士課程修了者が少ない場合，他の変数を調整しても博士課程進学者はほとんど変化しない」といった，単純な線形モデルでは説明できない構造を有する可能性が高い。こうした事例においてはむしろ，労働経済学において，線形モデルではなく指数型モデル（コブ・ダグラス型生産関数）が採用されることも参考に，「積の構造的因果モデル」[高山 22a, 高山 22b] を活用する方が適切であるとも考えられる。具体的には，式 (1) をもとに，以下のような指数モデルを仮定する。

$$x_i(t) = \prod_j \prod_{\tau=0}^{\tau_{max}} x_j(t-\tau)^{b_{i,j}(t,t-\tau)} e^{e_i(t)} \qquad (2)$$

今回活用するデータセットは全ての値が非負の値であることから，このモデルが適用可能である。また，式 (2) での探索のためのアルゴリズムを新たに開発する必要まではなく，原則的にはデータセットの各値について*10，全て自然対数をとった上で，式 (1) で表される Longitudinal LiNGAM のアルゴリズムを適用することで，積の構造的因果モデルによる探索としてを実行することが可能である。

---

*7 ただし，本モデルにおける注意点として，未観測共通要因がないと仮定しているということもあり，交絡を抑えた上で探索するために，$t=0$ における変数同士の因果効果は分析しない。しかしこのことは，同一時刻内の因果関係を因果効果を見られなくなることを意味するわけではなく，$t>0$ において $b_{i,j}(t,t)$ を見ればよい。

*8 同じ変数集合をとりながら，何らかの特徴に基づいてグループ分けされた複数のデータセットについて，変数間の因果的順序についてはグループを統合して探索の上決定しつつ，LASSO 回帰に基づいた枝刈り・因果係数の算出についてはデータセットごとに行うアルゴリズムである，Multi Group DirectLiNGAM[Shimizu12] を参考に，この制約を入れている。

*9 経時変化については，時刻 t，$t+1$，$t+2$ の 3 時点で分析することとする。

*10 ただしゼロの場合には，対数値の発散を防ぐための処置が必要である。そこで本研究では暫定的に，修士課程修了者がいない大学・年度については今回は分析対象から除くこととし，他の変数についてゼロがある場合は 1.0×10^{-5} に置き換えた上で対数処理を施して，因果探索を実行している。

■修士課程修了者数で場合分けした上での分析　アルゴリズムを適用するにあたって，大学による性質の違いを抽出しながら分析することを見据えると，何らかの基準で分類の上，別々に分析することも有効である。そこで各変数間の関係について，特に修士課程修了者数と運営費交付金の配分額についてプロットすると，図 1 に示す通り，強い正の相関がある一方，修士課程修了者数 1500 を境に傾きが異なることが分かった。そこで本研究では修士課程修了者数が 1,500 人未満の場合と（以下，便宜的に “under1500” と呼ぶ），1,500 人より大きい場合（“over1500” と呼ぶ）で分けて分析し，必要に応じて比較する。

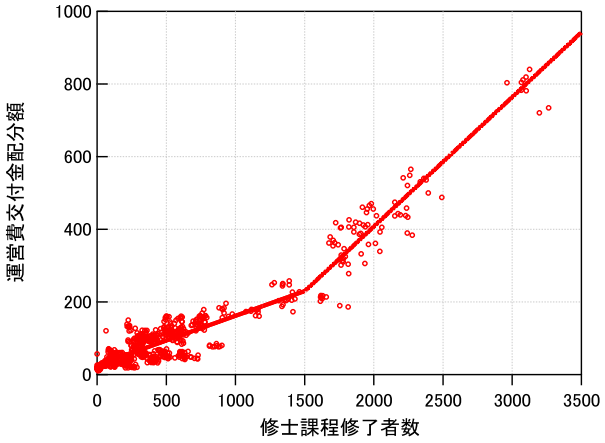


図 1: 修士課程修了者数と運営費交付金配分額の関係[*11]

■因果探索に当たっての事前知識　本アルゴリズムでは，他の LiNGAM 系アルゴリズムと同様，探索時の事前知識として，特定の変数間に因果関係がある/ないということを，係数計算時の要件として入れることができる。本研究では，以下の事前知識を入れて探索を実行することとした。

- 博士課程入学者数はあくまで修士課程修了者数の内数であり，修士課程修了者数の変化を引き起こすものではないため，$b^{(g)}_{0,1}(t,t) = 0$ とする。
- 修了者についても同様で，$b^{(g)}_{2,3}(t,t) = b^{(g)}_{2,4}(t,t) = 0$ とする。
- DC1 採択者数は毎年秋に確定するものであり，年度の頭より前に確定している他の変数に遅れることを踏まえ，修了者と同様に，$b^{(g)}_{7,j}(t,t) = 0$ とする。

[*11] 直線は修士課程修了者数 1,500 未満，1,500 以上でそれぞれフィットしたものであり，比例係数の推定値（± 標準誤差）はそれぞれ 0.136(±0.004)，0.356(±0.017) と推定された。

3　計算結果・考察

3.1　因果探索の結果の概要

まず，データセット全体で因果探索した結果として得られた DAG を図 2 に示した。ここでは DAG の解釈を容易にするため，自己回帰に該当する部分を図 2(b) に切り出し，それ以外の部分を図 2(a) に示している。全体を概観すると，under1500 と over1500 で，それぞれにしか現れていない edge が存在するなど，違いが散見される。

次に，この因果探索で得られた DAG の統計的信頼性を確認するため，図 2 の構造に基づいて，共分散構造分析[*13]を実行した。主要な結果を表 3 に示す。なお，この分析においても積の構造的因果モデルを仮定している。

表 3: 共分散構造分析による主な計算結果。

	#1	#2	CFI	TLI	RMSEA	AIC	BIC	#3
u1500	57本	50本(88%)	0.96	0.94	0.10	197.29	628.27	2.36
o1500	63本	56本(89%)	1.00	1.00	0.03	207.43	436.75	3.28

#1: 探索で現れたedgeの本数，**#2**: 係数値が有意なedgeの本数(割合)，**#3**: log-likelihood

まず各 edge の回帰分析の係数値の有意性という観点からは, under1500 でも over1500 でも, 有意水準 $\alpha = 5\%$ とした際，Multi Group Longitudinal LiNGAM での探索で出力された edge のうち，有意な edge は 85% 以上を占めている。また, 例えば CFI や TLI は 0.9 以上, RMSEA は 0.1 以下となっており，これらの構造スコアから見ても悪くないフィットである。

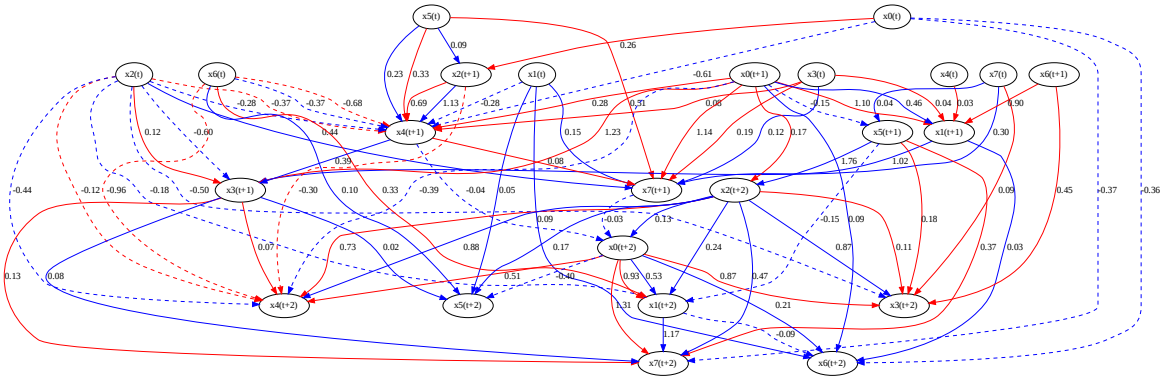
なお，以降解釈を行うにあたり，図 2(b) に示される通り，遅れを伴って表れている自己回帰に該当する edge が，探索結果の少なくない割合を占めている。この点は今後重要な論点になり得る一方，自己回帰の解釈は，変数によっても異なる可能性が高く，既存の領域知識からの議論も複雑なものとなることが想定されるため，本稿では個々には立ち入らず，以降は 2(a) に示されるような自己回帰以外のパスに絞って考える。

3.2　edge ごとの特徴

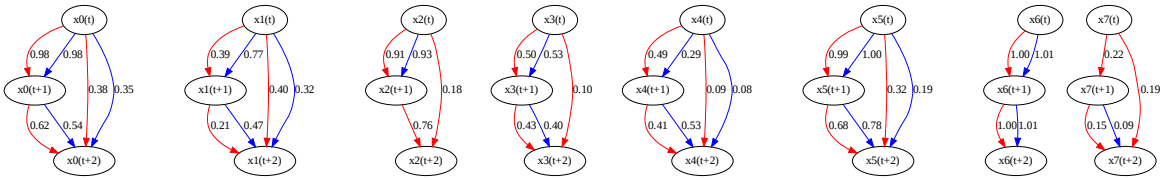
次に，edge ごとの特徴について，特に under1500 と over1500 を対比しつつ議論する。データセット全体で因果探索をして得られた図 2(a) の edge の中でも特に尤もらしい edge を抽出するため、bootstrap sampling を行った。図 3 には，その中でも bootstrap 確率が高く，はっ

[*12] 赤矢印は $x_0(t)$ が under1500 の edge，青矢印は $x_0(t)$ が over1500 の edge。また係数が正の場合を実線，負の場合を点線で示している。その他，これらの DAG に記載している edge は係数の絶対値が 0.01 以上のものに限定している。

[*13] 本稿では Python の共分散構造分析用パッケージである semopy(`https://semopy.com/`) を使用している。



(a) 探索結果のうち自己回帰に該当するパスを除いたもの。



(b) 探索結果のうち自己回帰にあたる部分のみを抽出して並べたもの。

図 2: データセット全体での Multi Group Longitudinal LiNGAM での探索結果*[12]

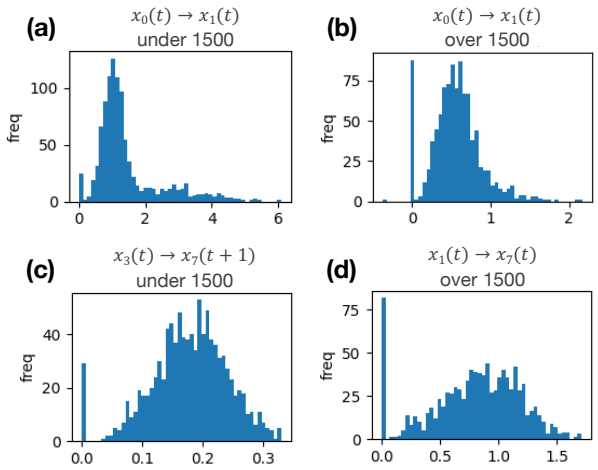


図 3: 特徴的な bootstrap ヒストグラム。

きりとした分布を示し，かつ特徴的なものについてのヒストグラムを並べている。 まず，under1500 と over1500 で共通する edge の一つに，$x_0(t) \to x_1(t)$ があり，それぞれ図 3 の (a)・(b) に示している。これは，先述の通り $x_1(t)$ が $x_0(t)$ の内数となっていることを考えると自然な結果ではある。一方で，この係数値については，仮に 1 であれば，博士課程進学率が一定であるという意味になるが，under1500 での分布の中心は付近であるのに対し，over では 1 を下回る。これは，edge としては共通でありながら，under1500 と over1500 との間に表れている決定的な違いの一つである。この背景としては，大学院博士課程の定員充足率の違いなどが予想されるが，修士課程修了者数を増やしても，博士課程修了者は，特に over1500 では増えにくいという傾向が，このような多変量解析でも明確に示唆されたことになる。

次に，DC1 採択者数への影響の出方の代表的な違いとして，図 3 の (c) に under1500 での $x_3(t) \to x_7(t+1)$ を，(d) に over1500 での $x_1(t) \to x_7(t)$ を示している。前者は，over1500 では現れにくく，後者は under1500 でほとんど現れない edge である。このような，DC1 採択者数への直接の効果の差異の起源として，DC1 採択に向けた研究・申請書執筆指導の在り方が under1500 と over1500 で異なることが考えられる。例えば，over1500 はそもそも博士進学者も多いので，先輩の申請書作成のノウハウを，1 年上の先輩からもらえることも多いと思われる。一方 under1500 では，1 年上の先輩からの指導よりも，学振 PD を取得したポストドクターからのノウハウ提供や研究指導による影響が大きい可能性がある。今後この点を深堀するには，例えば OpenAlex のデータを活用して，大学別の年間の論文数などを見ていくなど，研究成果の創出状況と一体的に見ていくことも有効であると考えられる。

3.3 直接・間接を問わない総合的な寄与

最後に，特に博士課程進学者数 $x_1(t)$ への他の変数からの総合的な因果効果 (total causal effect, TCE) と bootstrap 確率についても計算し，表 4 にまとめている。主な特徴は以下の通りである。

表 4: 博士課程進学者数に 1 年もしくは 2 年のラグを伴って表れる，各変数からの total causal effect と bootstrap 確率。

from	to	TCE		Probability	
		u1500	o1500	u1500	o1500
$x_0(t)$	$x_1(t+1)$	2.25	2.09	1.000	1.000
$x_2(t)$	$x_1(t+1)$	0.35	0.93	1.000	1.000
$x_3(t)$	$x_1(t+1)$	0.29	0.27	1.000	0.352
$x_4(t)$	$x_1(t+1)$	0.34	0.60	1.000	1.000
$x_5(t)$	$x_1(t+1)$	2.39	0.97	1.000	1.000
$x_6(t)$	$x_1(t+1)$	2.07	-1.07	0.937	0.970
$x_7(t)$	$x_1(t+1)$	0.27	0.57	1.000	1.000
$x_0(t)$	$x_1(t+2)$	2.31	2.30	1.000	1.000
$x_2(t)$	$x_1(t+2)$	0.32	0.99	1.000	1.000
$x_3(t)$	$x_1(t+2)$	0.27	0.64	1.000	0.618
$x_4(t)$	$x_1(t+2)$	0.34	0.61	1.000	0.787
$x_5(t)$	$x_1(t+2)$	2.45	0.64	1.000	1.000
$x_6(t)$	$x_1(t+2)$	1.85	-1.11	0.895	0.359
$x_7(t)$	$x_1(t+2)$	0.26	0.64	1.000	1.000

TCE: Total Causal Effect

- 博士課程進学に影響する要因として選んだ $x_1(t)$ 以外のほとんどの変数からの影響は，under1500 でも over1500 でも，1 年もしくは 2 年のラグを経て現れるという結果となっている。
- 全体的に over1500 の方が under1500 よりも TCE の絶対値は大きめだが，under1500 では $x_5(t)$(運営費交付金の配分額) の影響が over1500 よりも大きくなっている。
- $x_6(t)$(教員一人当たり学生数) による効果は，over1500 では負だが under1500 では正となっている。この符号の逆転についても，under1500 と over1500 での研究環境の構造的な問題と関係している可能性がある。

4 おわりに

本研究では，構造的にデータ量が確保しづらいという，「博士課程進学」に係わる統計的因果探索の課題に対応するため，

1. 国立大学の大学別データセットを構築することで，データ点数を増やし，
2. 更に統計的因果探索アルゴリズムについても改良の上，時系列を考慮した分析を行った。

結果として，ある程度妥当と思われる因果構造が計算され，さらに，修士課程修了者 1500 人未満と 1500 人超えの場合で，DC1 採択者数への直接の影響の表れ方や，博士課程進学者数への他の変数からの total causal effect について，それぞれ異なる傾向を観察できた。これらの傾向は，大学の研究環境の違いなどを反映している可能性もあり，今後高等教育論の観点からも詳細な解釈がなされることで，博士課程進学に係る定量的な因果構造の解明にさらに一歩近づくことが期待される。

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

[Hyvarinen10] A. Hyvärinen, K. Zhang, S. Shimizu, and P. O. Hoyer, Estimation of a structural vector autoregression model using non-gaussianity, *Journal of Machine Learning Research*, *11: 1709–1731*, 2010. https://www.jmlr.org/papers/volume11/hyvarinen10a/hyvarinen10a.pdf

[Hyvarinen13] A. Hyvärinen and S. M. Smith.： Pairwise likelihood ratios for estimation of non-Gaussian structural equation models. *Journal of Machine Learning Research*, 14:111–152, 2013. https://jmlr.org/papers/v14/hyvarinen13a.html

[Komatsu10] Yusuke Komatsu, Shohei Shimizu, and Hidetoshi Shimodaira：Assessing statistical reliability of LiNGAM via multiscale bootstrap. *In Proc. International Conference on Artificial Neural Networks (ICANN2010), Thessaloniki, Greece*, pp.309–314, 2010. https://doi.org/10.1007/978-3-642-15825-4_40

[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

[治部 21] 治部眞里, 星野利彦： 修士課程（6 年制学科を含む）在籍者を起点とした追跡調査（2020 年度修了（卒業）者及び修了（卒業）予定者に関する報告). **文部科学省 科学技術・学術政策研究所 調査資料** *(Research Material)*, No.310, 2021. https://doi.org/10.15108/rm310

[Shimizu12] Shohei Shimizu：Joint estimation of linear non-Gaussian acyclic models. *Neurocomputing*, 81:104-107, 2012. https://doi.org/10.1016/j.neucom.2011.11.005

[Kadowaki13] K. Kadowaki, S. Shimizu, and T. Washio： Estimation of causal structures in longitudinal data using non-Gaussianity. *2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, Southampton, UK, pp.1-6, 2013. doi:10.1109/MLSP.2013.6661912

[高山 22a] 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 22b] 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： 統計的因果探索アルゴリズム “LiNGAM” を活用した専攻分野別の博士課程進学に関する研究. **研究イノベーション学会 第 *37* 回年次学術大会（予稿集）**, 公演番号 1C06, 2022.

[高山 22c] 高山正行, 小柴等, 川村真理： ベイジアンネットワークによる博士課程進学要因の試行的分析. **研究イノベーション学会 第 *37* 回年次学術大会（予稿集）**, 公演番号 1C07, 2022.