

論文審査の要旨及び担当者

No.1

報告番号	甲 第 号	氏 名	篠田 和彦
論文審査担当者	主 査	： 長倉 大輔	(慶應義塾大学経済学部教授、Ph.D.)
	副 査	： 片山 翔太	(慶應義塾大学経済学部准教授、博士 (理学))
		： 星野 崇宏	(慶應義塾大学経済学部教授、博士 (経済学))
	面接担当	： 沖本 竜義	(慶應義塾大学経済学部教授、Ph.D.)
		： 中妻 照雄	(慶應義塾大学経済学部教授、Ph.D.)
(論文審査の要旨)			
論文題目 : Essays on Estimation of Causal Parameters Identified as the Ratio of Conditional Expectation Functions			
1. 論文の概要			
<p>本論文は、条件付き期待値関数の比 (Conditional Expectation Functions Ratio: CEFR) として識別される因果パラメータの推定法についての研究をまとめたものである。通常、処置効果としては潜在的結果の期待値の差として定義される Conditional Average Treatment Effect (CATE)などが検討されるが、問題によってはリスク比などの潜在的結果の期待値の比として定義される処置効果を考える方が自然な場合がある。さらに本論文の2章および3章で論じられるように、差で定義される処置効果であっても、データ融合の問題設定においてはCEFRとして識別されることがある。</p> <p>1章では、潜在的結果アプローチ、操作変数など、社会科学における因果推論の基本的な概念を導入し、CEFRとして識別される因果パラメータの例として、比で定義される処置効果、データ融合による処置効果推定などを挙げる。また、本論文の各研究に関連するトピックの文献レビューも行っている。</p> <p>2章はデータ融合によるLocal Average Treatment Effect (LATE)の識別および推定を扱う。処置割当に従わない個人がいる場合、実際に受ける処置は自己選択によって決定されるため、Unconfoundednes (潜在的結果と実際の処置の独立性)を前提とした因果推論は妥当ではない。このような場合でも、処置割当を実際の処置の操作変数として用いることで、割当に常に従う人たちに対する平均処置効果であるLATEが識別できることが知られており、これまでに多くの研究蓄積がある。既存のLATE推定法の多くでは、結果と割当、実際の処置と割当が同一のユニットについて同時に観測される必要があったが、実際には技術的あるいはプライバシー上の理由から、そのようなデータを集めることが困難か不可能な場合がある。本研究の貢献の一つは、処置割当確率の異なる2つのデータセットを用いることで、結果と実際の処置が同時に観測されず、処置割当が全く観測できない場合でも、LATEがノンパラメトリック識別できることを示した点である。さらに識別条件を整理することで、本章の設定の下では通常の処置効果推定で必要な正値性条件が不要となることを指摘し、このことを利用した実用的な実験・データ収集デザインを提案している。</p>			

論文審査の要旨

No. 2

推定に関しては、データ融合による CATE 推定の先行研究で提案された Direct Least Squares (DLS) が LATE 推定にも適用可能であることを示した。しかし、DLS はミニマックス問題の解として定義されるため、単純に目的関数を最小化するモデルが最良とは限らず、モデル選択が不安定になるおそれがある。そこで本研究はミニマックス問題を回避し、よりシンプルなモデル選択を可能とする推定量を提案しており、これも主たる貢献の一つである。さらに提案法は因果推論で広く用いられている逆確率重み付き推定法(IPW)とは異なり、事前に推定した傾向スコアを逆数としてではなくそのまま利用するため、IPW が引き起こす問題も回避することが可能である。シミュレーションでは、複数の関数形や共変量次元の設定において、既存手法と比較して提案法が一貫して精度が高かった他、IPW の問題点に対する頑健性も確認された。また、アメリカの職業訓練 (Job Training Partnership Act) データに提案法を適用し、先行研究と整合的な結果であった。

3章では、2章や4章の問題も含む一般的な CEFR 推定・推論の問題に対して、関数形の仮定を一切必要としない新しい枠組みを提案する。既存研究では CEFR の中でもリスク比や LATE などの限定された対象を検討しているが、本研究ではそれらを統一的に扱える点に新規性、有用性がある。提案の枠組みは二段階の方法であり、第一段階として Debiased Machine Learning (DML) を応用して Neyman Orthogonality と呼ばれるある種の直交性を満たすシグナルを構成する。DML とは結果の回帰関数や傾向スコアといった局外パラメータの推定に機械学習を用いても、目的の推定量への影響が漸近的には無視できるようにする技術である。社会科学分野では観察研究によって得られたデータの分析が多く、内在する選択バイアスを補正するために局外パラメータの推定量をプラグインする必要があるが、これらの局外パラメータは非常に高次元で非線形性が高い可能性があり、DML によって柔軟な機械学習を許容することで提案手法の実用性を高めることができる。第二段階では、第一段階で構成したシグナルを CEFR 推定用に新たに開発したシリーズ推定量と組み合わせることで、最終的な推定結果を得る。このシリーズ推定量は、データがランダム化比較試験から収集された場合などでは単体で用いることができ、これ自体本研究の貢献の一つと言える。さらに本研究では、シリーズ推定や DML に関する既存研究と同様の (抽象的な) 仮定の下で各点および一様の収束レートや漸近正規性を導出し、いくつかの具体例で局外パラメータ推定のレートに関する十分条件を示した。シミュレーションでは、ほとんどの場合において提案法は高精度の推定を実現するものの、非常にまれにかなり外れた推定となってしまうことがわかった。これはシグナルの構造に起因すると考えられ、この改善は今後の課題である。またアメリカの 401(k) データを用いた LATE を推定に提案法を適用した。

4章では、治癒モデルにおける処置効果推定の方法を提案する。治癒モデルとは、通常の生存時間分析とは異なり、ある割合の個人は治癒している、つまりどれほど時間が経過しても対象のイベントを決して経験しない集団が存在することを仮定したモデルである。様々な分野で治癒割合は重要な意味を持つが、治癒モデルのノンパラメトリック識別には通常、**Sufficient Follow-up**（追跡期間が十分長い）という強い仮定が必要である。しかし、実際には実験リソースの制約から十分な追跡期間を確保できず、治癒モデルを適用した分析が困難な場合がある。本章では、たとえ追跡期間が十分でなくても、処置群と対照群における非治癒割合の比をノンパラメトリックに識別できることを示した。非治癒割合の比は処置が治癒確率に与える影響であるから、処置の因果効果と解釈することができる。さらに、識別結果を **CEFR** に変形することで第3章の枠組を適用できることを示す。シミュレーションおよび実データを用いた実験により、提案法の性能を確認した。

2. 論文の評価

処置効果の推定に関する研究は計量経済学に限らず、様々な分野で行われており、これまでに膨大な蓄積があるが、最近では必要な変数が欠けていたり、従来用いられてきた強力な仮定を緩和した設定などの研究を通して、より実用的な因果推論手法の開発が進んでいる。このような情勢にあつて、本論文の主たる貢献は以下の2点に整理できる。1. データ融合や治癒モデルの文脈における、処置効果のノンパラメトリック識別に関する新たな結果を導出し、従来の知見では因果推論が不可能だった状況での選択肢を提示したこと、2. 広範な状況に適用可能な **CEFR** 推定方法を提案し、その漸近的性質を示したこと、である。以上の2点はどちらも現実的な仮定の下での因果推論の有力な方法を提案するものと評価できる。なお、本論文の2章の研究は人工知能研究分野のトップ国際会議である **AAAI** に採択されており、3章の内容のうち、より単純化された一部も三田学会雑誌に採択されている。さらに3章全体や4章の研究も現在投稿準備中である。

しかしながら、審査委員からはいくつかの疑問点や問題点も提起された。例えば、2章の設定では処置割当は観測不要という結果が示されたが、観測できる場合にそれを活用する方法についての議論が望まれる、3章の推定量の漸近的性質を導く際に推定対象の関数の有界性を仮定してしまっているが、もっと弱い条件があるのではないかと、**Neyman Orthogonality** を満たすシグナルは分析者が実際の推定対象ごとに用意する必要があるため、そこまで汎用的な方法とは言えないのではないかと、といった意見が挙がった。ただし、これらの指摘は本論文の学術的な価値や貢献を損なうものではなく、提案された諸手法の有用性をさらに高めるための今後の研究課題と言える。

以上の所見から、本論文は因果推論の分析の枠組みのさらなる拡張に資する新規性のある研究であり、審査委員会は全員一致で本論文が博士（経済学）の学位を授与するにふさわしい内容であると判断する。