

# 論文要旨

## 操作変数法に基づく因果効果推定における モデル選択に関する検討

生物統計情報学コース

49-226611

玉野 史結

観察研究における因果効果の推定では、未測定の交絡因子に由来するバイアスがしばしば問題になる。この問題に対処する手法の一つに操作変数法がある。操作変数法は、操作変数とよばれる特定の条件を満たす変数を用いる方法であり、曝露に対する操作変数のモデルとアウトカムに対する曝露のモデルの二段階のモデルに基づいて因果効果の推定方法を構築する。操作変数を含む調整変数集合の選択は背景知識に基づいて行われることが多いが、候補となる変数集合が複数存在する場合に、その選択が最終的な因果効果の推定量の性質にどのように寄与するかの知見は限定的である。また、データに基づくモデル選択手法の適用についても先行研究はほとんど存在しない。

本研究の目的は、操作変数法に基づく因果効果の推定において、調整変数集合の選択が最終的な因果効果推定における統計的性能にどのように影響するかを明らかにするとともに、情報量規準に基づくモデル選択を組み込んだ推定方法について、使用する情報量規準に応じた性能を比較することである。特に、複数の操作変数および共変量を用いる状況を想定し、操作変数法に基づく推定手法として代表的な 2-Stage Residual Inclusion (2SRI) と Limited Information Maximum Likelihood (LIML) を用いる状況を考察対象とした。シミュレーション実験によって、使用するモデルによる因果効果の推定性能の差異の確認、および、情報量規準によるモデル選択を組み込んだ推定方法の性能比較を行った。後者で使用する情報量規準としては、先行研究でも検討された赤池情報量規準 (AIC) およびベイズ情報量規準 (BIC) に加え、本研究の設定における FIC (focused information criterion) の具体形を導出

した。推定性能はバイアスと二乗平均平方根誤差(RMSE)を指標として比較した。

調整変数集合の選択による差異を調べたシミュレーション実験では、特に操作変数の条件を満たさない変数が含まれる設定において、調整変数集合の選択が最終的な因果効果の推定性能に大きく影響する様子が確認された。この設定の下で、情報量規準によるモデル選択を組み込んだ手法を比較した結果、FIC に基づくモデル選択を組み込んだ推定方法による因果効果の推定値に対するバイアス、RMSE が最も小さかった。一方、すべての操作変数の候補が条件を満たす設定においては、AIC や BIC に基づく方法の優位性が確認された。これらの結果から、操作変数の条件を満たさない変数の混入が疑われるかどうかによって、有用な情報量規準が異なる可能性が示唆された。以上は特定のモデル設定と限定的なシナリオでの実験結果に基づくことから、操作変数法におけるモデル選択手法についてさらなる検討が必要である。