

# 論文審査の要旨及び担当者

No.1

報告番号	甲 第 号	氏 名	加藤 諒
論文審査担当者	主 査	：中妻 照雄 (慶應義塾大学経済学部教授 Ph.D.)	
	副 査	：長倉 大輔 (慶應義塾大学経済学部教授 Ph.D.)	
		：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	：新井 拓児 (慶應義塾大学経済学部教授 博士(理学))	
		：河井 啓希 (慶應義塾大学経済学部教授 商学修士)	
(論文審査の要旨)			
論文題名 : Studies in Semiparametric Causal Inference and Missing Data Analysis			
1. 論文の概要			
<p>本論文の主たる目的は、因果推論と欠測データの問題に対して、セミパラメトリックモデルの開発と応用を行うことである。ある処置に対する因果効果の推定は、様々な学術分野で最大の目的となりうる。特にランダム化比較実験が行うことが困難である社会科学や自然科学においては、準実験としての潜在的結果変数フレームワークを用いた因果推論が行われるが、これは複数ある潜在的結果変数のうち一つのみ観測されそれ以外は欠測されるという点で統計学的には欠測データと同じ問題として議論できる。</p> <p>最もよく用いられる因果効果推定の方法の1つとして傾向スコア解析が挙げられるが、これは観測されない交絡要因が存在する場合には頑健な方法ではない。一方、操作変数法は適切な操作変数を利用することができれば、正しい因果効果を推定することが可能である。そのため、操作変数法は経済学だけではなく、医学・疫学などの分野においても近年広く用いられ始めている。その一方で、欠測データの問題とその対処に関しても、近年より多くの関心が寄せられている。近年の大規模データの利用可能性の拡大に伴い、多くの変数を解析に利用することができるが、変数を多く利用すればするほど、欠測を含むオブザベーションが増えることから、欠測データに対する解析法の開発へのニーズはますます高まってきている。</p> <p>本研究では、これらの操作変数法を含む因果推論と欠測データの問題に対して、より柔軟なモデルを開発し、応用することを目的としている。ここでは特にセミパラメトリックモデルを利用することで、現実の問題に対処可能な因果推論と欠測データ解析のための方法を提案・応用する。セミパラメトリックモデルにおいては、特に関心のある因果効果などのモデルにはパラメトリックモデルを仮定し、関心のないモデル、あるいは特定化が困難なモデルに対してはノンパラメトリックモデルを仮定する。これにより、より少ない仮定の下で関心のあるパラメータの推定を正しく行うことができる。</p> <p>本論文は6つの章から構成されている。第1章は導入、第6章は結語であるため、ここでは第2章から第5章の概要を手短かに紹介したい。</p>			

# 論文審査の要旨

No.2

第2章では、現実の社会科学におけるセミパラメトリックな因果効果推定の重要性について議論し、具体的な応用例として、傾向スコア法を用いた会計学における因果推論を行った例を紹介した。ここでは、監査法人の規模の違いが監査の質に与える影響に関する因果推論を行った。規模の大きな監査法人はより大規模な企業がクライアントになりやすく、結果としてクライアントの会計報告の質がそもそも高いため、監査法人の規模が監査の質に与える影響を知るためには、この点を調整する必要がある。そこで傾向スコア法を用いてクライアントの属性を事前に調整し、監査の質とされる裁量的発生高に対する因果効果の推定を行った。その結果、監査法人の規模を十分に調整していなかったこれまでの先行研究とは異なり、監査法人の規模の違いは監査の質に影響を与えないことが示された。

第3章では、欠測データに対処する方法として、セミパラメトリックベイズな多重代入法の提案を行った。欠測データを無視して解析を行った場合、誤った推論を行うことが示されているが、これに正しく対処する方法として多重代入法が提案されている。最もよく用いられる多重代入法として、連鎖式による多重代入法 (MICE; Multiple Imputation by Chained Equation) が広く用いられるが、これは特殊な状況を除いて一致性のある推定を行う保証がないことが近年示された。そこで本章では、変数間の関連性を示すモデルにはノンパラメトリックモデル、因果効果といった関心のあるパラメータにはパラメトリックモデルを仮定するセミパラメトリックのベイズ多重代入法を提案した。また潜在変数の考え方を用いることで、連続変数・離散変数が混合しているデータセットにも対処することができる方法である。シミュレーションの結果、MICE を含む既存の方法よりも、かなり広い状況においてよりよい推定量が得られることが示された。ただし第3章では、より仮定の強いランダムな欠測の状況を扱っているため、ランダムでない欠測への拡張が望まれる。この拡張は第5章で扱う。

第4章では、因果効果推定のための方法論として、セミパラメトリックなベイズ操作変数法を提案している。近年、遺伝子情報などが利用可能となり、完全に外生的に与えられると考えられるこの情報を操作変数として扱うメンデルアン・ランダムマイゼーションが医学・疫学や経済学で用いられるようになった。メンデルアン・ランダムマイゼーションは、経済学で用いられてきた操作変数と数理的に同じものであるが、操作変数の次元がより大きいという違いがあり、この点に対処できる操作変数法を開発する必要があった。そこで本章では、操作変数と内生変数の関係を記述する誘導方程式にはノンパラメトリックベイズモデル、内生変数と目的変数の関係を記述する構造方程式にはパラメトリックモデルを仮定する、セミパラメトリックベイズ操作変数法を提案した。提案手法は、応用研究でよく利用される既存の一般化モーメント法や制限情報最尤法から得られる推定量よりも効率性が高いことが示され、また実データへの応用からも理論的には予測されながらも既存の方法では得られなかった結果を得ることができた。

第5章では、第3章では扱わなかったランダムでない欠測を扱うことを目的として、第4章で扱った操作変数に関する欠測の問題に対処するセミパラメトリックモデルを提案した。操作変数は、その他の説明変数や目的変数と比べて、データソースが異なることが多く、操作変数の欠測が起りやすいという問題がある。これに対してランダムでない欠測を仮定する場合、欠測メカニズムをモデル化する必要があるが、一般にこのモデリングは困難であることに加え、モデル自体を識別できないという重要な問題がある。そこで第5章では、操作変数の母集団の分布を利用できるという仮定を置き、ノンパラメトリックな欠測メカニズムを仮定しても、パラメトリックな構造方程式が識別されることを示した。さらにここでは、欠測データの代入と、操作変数法における因果効果パラメータの推定を同時に行うマルコフ連鎖モンテカルロ法を提案した。シミュレーションを行った結果、既存のデータ代入法では一致性のある推論が行えないのに対して、提案手法は一致性があり、効率性も高い推定量が得られることが示された。また、提案法を法制度がGDPに与える影響を調べたAcemoglu et al. (2001)のデータセットに適用した。ここで操作変数の「入植者の死亡率」に欠測があるが、Albouy (2012)によって改良された死亡率データを母集団分布と仮定して提案方法を適用した場合、法制度とGDPには既存研究で示されたより強い関連があることが示唆された。

## 2. 論文の評価

近年のデータ利用可能性の拡大に伴い、欠測データ解析と因果推論の問題は多くの学術分野においてますます重要になっているが、経済学においても多様なデータを利用した実証が盛んになりつつある。このような環境の変化を踏まえると、本論文が提案するセミパラメトリックな因果推論・欠測データ解析法は、経済学のみならず、医学・疫学などを含む幅広い分野での調査観察研究における実証分析の重要なツールとなりうるであろう。なお、本論文の元となった研究成果は、3章の内容が統計学におけるQ1ジャーナルであるAnnals of the Institute of Statistical Mathematicsに採択されており、またほかの章の内容も複数の査読付き論文として英文雑誌において公刊済みあるいは投稿中である。

しかしながら、セミパラメトリックモデルの中の一部で母集団に関する仮定を置くことから、その仮定の妥当性の検証方法などについての議論や、セミパラメトリックモデルから得られる推定量のより詳細な統計的な性質に関する議論が望まれるという声も審査委員の中にはあった。

但し、上記の指摘された課題は本論文の学術的な価値や貢献を損なうものではなく、ここで提案された諸手法が幅広く利用されるための今後の研究課題と言える。

以上の所見から、本論文は計量経済学や幅広い調査観察研究への応用という点において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と判断する。