

論文審査の要旨及び担当者

No.1

報告番号	甲 第 号	氏 名	高畑 圭佑
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))
		： 片山 翔太	(慶應義塾大学経済学部准教授、博士 (理学))
(論文審査の要旨)			
論文題目			
Identification and estimation of the joint distribution of potential outcomes in causal inference			
1. 論文の概要			
<p>本論文は、因果推論における Potential Outcome (潜在的結果変数) の同時分布を対象として、その識別性および推定法についての議論を行っている。因果推論において通常研究の関心とされるのは、平均処置効果 (ATE: Average Treatment Effect) や処置群における平均処置効果 (ATT: ATE for the Treated) といった母集団の平均に関するパラメータである。これらのパラメータは Potential Outcome の周辺分布の関数であるため、分析の際に Potential Outcome の同時分布については議論を行う必要はない。</p> <p>一方、近年では、ATE や ATT といった母集団の平均に関するパラメータにとどまらず、因果効果の分布も考慮した議論を行う必要性が認識され始めている。具体例としては、i) Potential Outcome 間の相関、ii) (本来の定義に基づく)分位点処置効果 (QTE: Quantile Treatment Effect)、iii) 異質性のある処置効果 (HTE: Heterogeneous Treatment Effect) などの問題を挙げることができる。これらのパラメータの推論を行う際には、Potential Outcome の同時分布を評価する必要があるため、前提として同時分布の識別性を保証する必要がある。しかし、Potential Outcome は実現した処置に対応するもののみが観測されるため、全ての Potential Outcome が同時に観測されている個体は存在せず、一般的に観測されたデータのみでは同時分布を一意に定めることはできない。したがって、分布の関数形への制約、操作変数に相当する変数の存在の仮定など、何らかの追加条件が必要となる。</p> <p>以上のような背景の下、本論文では Potential Outcome の同時分布をパラメトリックに識別する方法、ノンパラメトリックに識別する方法をそれぞれ提案するとともに具体的な推定に関する議論を行っている。同時分布の情報の重要性は先行研究でも指摘されているものの、具体的には部分識別の枠組みで関連する議論</p>			

を行ったものが僅かに存在するのみである。本論文の特筆すべき貢献の1つは、同時分布について部分識別ではなく点識別のための十分条件を初めて与えたことにある。導出された条件は多くの応用上の問題に適用可能であり、因果効果推定の研究における重要な基礎研究の一つとなると考えられる。また、導出された識別性条件の下でのベイズ推定の提案も行っている。

第1章は計量経済学における識別問題をサーベイした導入部であるため、以下では本論文の主たる貢献を為す第2章から第4章の内容を要約する。

第2章では、同時分布をパラメトリックに特定化する場合の識別性の議論を行っている。まず簡単な例をあげると、2つの Potential Outcome が2変量正規分布に従う場合は、周辺分布も正規分布となり相関係数に関するパラメータは一意に定まらないため、識別することができない。この結果を積率母関数を用いて分析すると、周辺分布の積率母関数と条件付き分布の積率母関数が同じ関数形である場合には、パラメータが積率母関数内の変数について集約されてしまうことが識別不能である原因である。逆に、積率母関数の一意性から、上述のパラメータ数の減少は正規分布(または一点に縮退した分布)以外であれば起きない。以上のことから、条件付き分布が正規分布である場合には Potential Outcome の周辺分布が正規分布もしくは縮退分布以外であれば同時分布を識別可能であることを示した (Theorem 1)。さらに第2章では、同様の分析を同時分布が混合正規分布で特定化されている場合へも拡張している。任意の連続分布は特殊な場合を除いて混合正規分布の成分数を十分増やすことで精度良く近似することが可能であるため、混合正規分布における識別性の議論を行うことは当該分析の射程を大きく広げると考えられる。導出された結果 (Theorem 2 と Theorem 3) には混合正規分布の各成分間の相関に制約を加える必要があるなど一見厳しい条件も含まれるが、成分数を十分増やせば応用上問題のない水準の柔軟性があることが確認された。以上の識別の結果は、数値実験によっても支持された。

第3章では、第2章で提案された方法をさらに拡張し、複雑なサンプル・セレクションが存在する実データへ応用を行っている。具体的には、米国で行われた職業訓練プログラムの一つである、Job Training Partnership Act (JTPA) のデータを用いて、HTE のベイズ推定を行った。JTPA データの特徴は、大規模なランダム化比較試験 (RCT: Randomized Controlled Trial) である一方で、処置群の方にのみ無視できない割合 (約40%) の不遵守 (One-sided Noncompliance) が存在することである。加えてその不遵守は条件付き独立性 (Unconfoundedness) の仮定を満たさない、すなわち欠測がランダムではない (MNAR: Missing Not At Random) ことが示唆されている。このことは、不遵守によるサンプルの歪みを考慮しつつ、同時分布の推定を行う必要があることを示している。第2章では、主に RCT の場合の識別性の結果を示したが、本章ではその結果を MNAR による欠測が存在する場合にも拡張している (Theorem 5)。加えて HTE の具体的な推定方法として、MNAR の欠測を調整するために、外部情報として対照群の積率情

報を GMM を活用して尤度に反映させる方法を提案している。これにより、JTPA と比較して得られる情報がより限定的であるデータに対しても提案手法は適用可能となる。女性労働者に対して行った分析では、対照条件下では、収入なし、あるいは低収入の人に対しては大きな処置効果が見込める一方で、中程度以上の収入である人に対しては処置効果が比較的低いことがわかった。これらの結果は今後の政策立案に対して示唆を与えるものと考えられる。

第 4 章では、同時分布をノンパラメトリックに扱う場合の識別条件の議論を行っている。周辺分布と同時分布は積分を通じて関連付けられるため、解くべき問題はある種の積分方程式の一意性に帰着する。ただし、当該問題における積分方程式は、一般的に積分方程式の文脈で考察の対象となる Fredholm 型の積分方程式とは形状が異なるため、そのまま理論的な分析を行うことは困難である。そこで Control Variable の仮定を導入することを提案している。Control Variable とは操作変数と類似の概念であり、労働経済学や医療経済学の実証においても同様の仮定が採用されている。当該積分方程式を共変量での条件付きで考えると、Control Variable の仮定の下では既述の Fredholm 型積分方程式の一種と解釈し直すことができる。したがって、この形に対して既存の積分方程式の一意性の結果（特に Volterra 型積分方程式）の応用することで、識別性を示すことができる。Volterra 型積分方程式の解の一意性は、積分範囲が関数の変数に依存していることに基づく。この点を考慮し、いくつかの正則条件に加えて、Potential Outcome のサポートが Control Variable に依存することを要求する Moving Support Condition が満たされれば、同時分布はノンパラメトリックに識別されることを示した (Theorem 6)。導出された識別条件は一見制限的であるが、本論文では応用上用いられるいくつかのモデルにおいて近似的に満たされることを数値実験によって確認している。

この理論的結果に加えて、第 4 章ではノンパラメトリックでの識別という特性を活かすため、Dirichlet Process Mixture Models に基づく推定方法を提案している。Potential Outcome の枠組みでは必ず観測されない変数が存在するため、観測データに関する尤度は積分を含み、その計算コストはサンプルサイズに比例して増大する。一方でベイズ統計学の枠組みではデータ拡大アルゴリズムを採用することにより、この積分を直接評価することなくマルコフ連鎖モンテカルロ法で推定することができる。本論文は、各個体がどの潜在クラスに属するかを示すインデックスを導入し、このインデックスや潜在クラス内のパラメータなどをギブス・サンプリングで生成するという効率的アルゴリズムを提案している。この提案手法を用いて数値実験を行ったところ、識別可能となったモデルに対しては正しく推定できた一方で、識別不能となったモデルの一部についても正しく推定できるという結果になった。上記の実験結果は、今回提案した識別条件が十分条件にすぎず、より弱い識別条件が存在することを示唆している。

2. 論文の評価

因果推論の文脈での平均処置効果に関する研究知見は、計量経済学に限定せず、数理統計学や医療統計学、計量政治学など様々な分野で膨大な研究蓄積があるが、近年では効果の平均にとどまらず、分布や異質性を考慮する取り組みが盛んに行われている。概要部分の冒頭でも述べた通り、Potential Outcome の同時分布の情報はこれらのような研究関心において数々の応用が期待されるが、本論文はその発展を考える際の一つの基礎を与えるものと評価できる。なお、本論文の元となった研究成果の中で、第 2 章の内容は International Statistical Institute が発刊する査読付きジャーナルである Stat に採択されており、第 3 章や第 4 章も現在投稿準備中である。

しかしながら、本論文はいくつかの問題点も内包している。例えば、コピュラを始めとする、周辺分布と同時分布を関連付ける既存理論との相違点についてより明確化すべきではないかという議論や、点識別の条件を導くために今回利用された積分方程式の解の一意性についての議論を踏まえて、必要十分に近いより弱い条件の提示が望まれるという指摘が審査委員の中から挙がった。

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、それらは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ることから、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。