

Covariate Selection for Estimating Population
Means and Procedures for Estimating Treatment
Effects When Data Are Missing Not at Random

January 2025

Shintaro Yoneyama

主 論 文 要 旨

No.1

報告番号	① 乙 第	号	氏 名	米山 慎太郎
主 論 文 題 名 : Covariate Selection for Estimating Population Means and Procedures for Estimating Treatment Effects When Data Are Missing Not at Random (Missing Not at Random データにおける母平均推定の共変量選択と処置効果の推定手順)				
(内容の要旨) 本論文は、結果変数に欠測がありそのメカニズムが Missing Not at Random (MNAR) の場合において、結果変数の母平均をバイアスと分散を小さく推定するための共変量選択、及び、処置効果を推定する方法について研究成果をまとめたものである。MNAR は観測値のみならず欠測値自身にも依存する欠測メカニズムであるが、その他の欠測メカニズムと比べ適切な調整が難しい。 第一章は序論であり、欠測データの調整法における先行研究について概観し、現状の MNAR データにおける統計的推測の問題意識について述べる。 第二章では、本論文での議論に用いる先行研究について説明する。ここでは、結果変数に MNAR の欠測がある場合の母平均推定において、共変量を条件付けたときに結果変数と条件付き独立で欠測とは関連がある変数である操作変数を用いた方法を考える。当該の先行研究は 2 段階推定手順を採用しており、まず共変量、結果変数及び操作変数を条件付けたもとの欠測確率である拡張傾向スコアを推定し、その推定値を用いて母平均の推定を行う。 第三章では、第二章で説明した 2 段階推定手順による母平均推定において、拡張傾向スコアのモデルに対する共変量選択指針の提案を行う。まず、母平均推定の一致性のために拡張傾向スコアのモデルに最小限含めるべき共変量を示す。次に、母平均推定値のバイアスと分散を小さく推定するという観点で、一致性のために最小限含めるべき共変量のみを拡張傾向スコアモデルに含めることが良い共変量選択指針であることを示す。 第四章では、結果変数に MNAR の欠測がある場合に 2 段階推定手順を用いて処置効果を推定する方法を提案する。観察研究において処置効果を推定する際は交絡変数の存在が問題になるので、欠測と交絡の調整を同時に考える必要がある。ここでは観察研究において結果変数に MNAR の欠測がある場合に、結果変数のモデリングを行うことなく、処置効果として平均処置効果や処置群・対照群での平均処置効果を推定する方法について提案する。 第五章は、結果変数に MNAR の欠測がある場合における母平均推定のための共変量選択、及び処置効果の推定手順についての議論をまとめる。また、MNAR を調整する 2 段階推定手順における課題と今後の展望についても述べる。				

Thesis Abstract

No. _____

Registration Number	☒ "KOU" ☐ "OTSU" No. _____ *Office use only	Name	Shintaro Yoneyama
Thesis Title Covariate Selection for Estimating Population Means and Procedures for Estimating Treatment Effects When Data Are Missing Not at Random			
Thesis Summary This thesis summarizes the results of research on covariate selection for estimating the population mean of the outcome with small bias and variance, and on estimating the treatment effects when the outcome is missing and the mechanism is Missing Not at Random (MNAR). MNAR is a missing mechanism that depends not only on the observed values but also on the missing values themselves and is more difficult to adjust appropriately than other missing mechanisms. This thesis is organized as follows. Chapter 1 is an introduction, reviewing previous studies on missing data adjustment methods and discussing the limitation of recent MNAR data adjusting methods. Chapter 2 describes the previous study used for the discussion in this thesis. In this chapter, we consider the use of instrumental variable (IV), which is the variable that is conditionally independent of the outcome and related to the missing mechanism when conditioning on the covariates, in the estimation of the population mean when the outcome is MNAR. The previous study employs a two-step estimation procedure, first estimating the extended propensity score, which is the probability of missingness conditional on the covariates, the outcome, and the IV, and then estimating the population mean using the estimates. In Chapter 3, we propose a policy for covariate selection for the extended propensity score model in the population mean estimation using the two-step estimation procedure described in Chapter 2. First, we show the minimum set of covariates that should be included in the extended propensity score model for the consistency of the population mean estimation. Next, we show that it is a good policy to include only the minimum set of covariates for consistency in the extended propensity score model in terms of estimating a small bias and variance of the population mean estimate. In Chapter 4, we propose a method to estimate treatment effects using a two-step estimation procedure when the outcome is MNAR. Since the existence of the confounder introduce bias when estimating treatment effects in observational studies, it is necessary to simultaneously consider adjustments for missing and confounding. In this section, we propose a method for estimating the average treatment effect, the average treatment effect on the treatment and on the untreated without modeling the outcome in the observational study where the outcome is MNAR. Chapter 5 summarizes the discussion on the covariate selection for estimating the population mean and the procedure for estimating the treatment effects when the outcome is MNAR. It also discusses the challenges of the two-stage estimation procedure for adjusting the MNAR and future work.			