別表５
⑶

主　　論　　文　　要　　旨

No.1

報告番号	甲　乙　第　　　　　号	氏　名	篠田和彦

主 論 文 題 名：
Essays on the Estimation of Causal Parameters Identified as the Ratio of Conditional Expectation Functions （条件付き期待値関数の比として識別される因果パラメータの推定に関する小論）

（内容の要旨）
This thesis consists of three essays on estimation of causal parameters that are identified as the ratio of conditional expectation functions (CEFR), and introduction to the problem of CEFR estimation in causal inference.

<u>Introduction (Chapter 1)</u>
In this chapter, I introduce basic concepts in causal inference in social sciences such as the potential outcome approach, unconfoundedness assumption, and instrumental variable estimation. Then, I provide some examples of the causal estimands that are identified as CEFR including ratio-based treatment effects and treatment effects in data combination settings. Literature review on the topics related to the studies in this thesis are also provided in this chapter.

<u>Estimation of Local Average Treatment Effect by Data Combination (Chapter 2)</u>
It is important to estimate the local average treatment effect (LATE) when compliance with a treatment assignment is incomplete. The previously proposed methods for LATE estimation required all relevant variables to be jointly observed in a single dataset; however, it is sometimes difficult or even impossible to collect such data in many real-world problems for technical or privacy reasons. We consider a novel problem setting in which LATE, as a function of covariates, is nonparametrically identified from the combination of separately observed datasets. For estimation, we show that the direct least squares method, which was originally developed for estimating the average treatment effect under complete compliance, is applicable to our setting. However, model selection and hyperparameter tuning for the direct least squares estimator can be unstable in practice since it is defined as a solution to the minimax problem. We then propose a weighted least squares estimator that enables simpler model selection by avoiding the minimax objective

主　論　文　要　旨

formulation. Unlike the inverse probability weighted (IPW) estimator, the proposed estimator directly uses the pre-estimated weight without inversion, avoiding the problems caused by the IPW methods. We demonstrate the effectiveness of our method through experiments using synthetic and real-world datasets.

Orthogonal Series Estimation for the Ratio of Conditional Expectation Functions (Chapter 3)

In various fields of data science, researchers are often interested in estimating the ratio of conditional expectation functions (CEFR). Specifically in causal inference problems, it is sometimes natural to consider ratio-based treatment effects, such as odds ratios and hazard ratios, and even difference-based treatment effects are identified as CEFR in some empirically relevant settings. This chapter develops the general framework for estimation and inference on CEFR, which allows the use of flexible machine learning for infinite-dimensional nuisance parameters. In the first-stage of the framework, the orthogonal signals are constructed by using the techniques of the debiased machine learning to mitigate the negative impacts of the regularization bias in the nuisance estimates on the target estimates. The signals are then combined with a novel series estimator tailored for CEFR. I derive the pointwise and uniform asymptotic results for estimation and inference on CEFR, including the validity of Gaussian bootstrap, and provide low-level sufficient conditions to apply the proposed framework to some specific examples. I demonstrate the finite-sample performance of the series estimator constructed under the proposed framework by numerical simulations. Finally, I apply the proposed method to estimation of the causal effect of the 401(k) program on the household assets.

Estimation of a Treatment Effect from Survival Data with a Cure Fraction Under Insufficient Follow-Up (Chapter 4)

I consider nonparametric estimation of a treatment effect from survival data while explicitly assuming that a certain proportion of subjects are cured, namely they never experience the event of interest, departing from the classical survival analysis. Although modeling a cure fraction is highly important as it can be found in a variety of research fields, it is sometimes hard to apply cure models in practice

主　　論　　文　　要　　旨

since nonparametric identification of cure models usually requires a strong assumption called sufficient follow-up. This chapter shows that the ratio of uncured subjects in treatment and control groups can be identified even if the follow-up period is not long enough, by avoiding identifying a cure model itself. Furthermore, I propose a novel efficient estimation method by applying the inference framework developed in Chapter 3. I illustrate the performance of the proposed method by numerical simulations. I also apply the proposed method to estimate the causal effect of the type of release on the recidivism rate in the US.