Université franco-allemande Deutsch-Französische Hochschule

Summary: Essays on the Econometrics of Program Evaluation: On the Reliability of Observational Methods Based on Unconfoundedness

Jasmin Fliegner

This dissertation aims at understanding and improving the reliability of observational methods.

The first two chapters of this dissertation focus on assessing the quality of observational estimation methods based on the unconfoundedness assumption, i.e. the assumption that the treatment is as good as random after controlling for the relevant covariates. In the first chapter, based on joint work with Sylvain Chabé-Ferret and Roland Rathelot, we propose a framework to estimate the bias of observational methods and understand its driving forces using a wide range of randomized controlled trials (RCTs) with imperfect but known compliance. Contrary to most earlier attempts, our method does not require the collection of additional data on non-participants and does not suffer from a bias resulting from using different survey instruments for participants and non-participants. Our proposed approach can also accommodate encouragement designs with treated individuals in the control group. We therefore propose a self-contained methodology applicable to a broad set of different RCTs which opens the possibility of collecting a multitude of biases from observational methods and understanding the driving forces of this bias. We also introduce a new decomposition of the bias of observational methods into two components: one due to unobserved confounders and the other due to lack of common support.

In the second chapter, as a proof of concept, we run this decomposition on publicly-available data corresponding to six published papers. We evaluate programs, mostly from developing countries, in education, labor, micro-finance and health using both a simple approach conditioning linearly on all observed confounders, and an approach using machine learning to select the relevant covariates. Using local linear kernel matching estimation, we find that in most cases the bias after conditioning on the observed covariates available in the datasets is as large as the bias before conditioning on anything. Our results suggest that the covariates we observe in this context are poor predictors of selection bias. We also find that the second component of the bias of observational methods due to a failure of common support is generally small.

For the future, we hope to have proposed a framework which permits additional datasets to easily be added to enrich the analysis and to investigate which confounders matter most for a given type of program and outcome.

In the first two chapters, we focus on the identification of the causal effect of a program. However, it is equally crucial to estimate correct standard errors and assess the corresponding large sample distribution in order to draw conclusions about the statistical significance of the treatment effect. In the third chapter of this thesis, I therefore propose an inference procedure for imputation-based matching estimators. This method is applicable to the class of estimators that impute the missing potential outcome as a weighted sum of outcomes from the opposite treatment group. Generalizing the methodology suggested by Abadie and Imbens (2006) for nearest neighbor matching estimators, I establish root-n-asymptotic normality of the matching estimator for the population average treatment effect minus a bias term. Moreover, I propose a generalized estimator of the corresponding marginal variance and derive a large sample variance estimator for the matching estimator for the population average treatment effect on the treated. This versatile way of estimating the standard errors is a robust and computationally more efficient alternative to the naïve bootstrap should the latter be valid. In a Monte Carlo study, I assess the performance of the proposed estimator for the marginal variance by using

a local linear matching estimator for the population average treatment effect on the treated. I obtain precise standard errors and coverage rates that perform equally well if not better than the naive bootstrap.

References

Abadie, A. and G. W. Imbens (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74, 235–267.