Research Site Selection Bias: Evidence from the Peruvian Amazon*

Yuma Noritomo†

January 20, 2021 [Latest ver. Available]

Abstract

Out-of-sample prediction may fail intuitively if one chooses study sites purposively due to research feasibility but the evidence on its consequences is scarce. By analyzing the first large-scale detailed surveys in the remote Peruvian Amazon context, this study tests for the research site selection bias specifically using studies that assess the impact of floods. I predict the average treatment effects (ATEs) in the unresearched communities identified by a systematic review based on the other research-available communities. The results are summarized as follows: First, the impacts are heterogeneous and significantly adverse for households in unresearched communities. Second, adjusting the covariate distribution of geographic variation improves the prediction of ATEs. Third, the predictions cause a false-zero error, possibly leading to wrong policy implications. As a mechanism, studying the selection of researched communities reveals that access from the cities and NGOs presence are systematically different. In this context, the formers lack of overlapping distribution reduces the preciseness of the prediction, while the latter does not induce a significant bias. The results highlight the difficulty of precise out-of-sample predictions given the research feasibility constraints in remote areas.

*I thank Yoshito Takasaki for the discussion. I appreciate Kazushi Takahashi, Eric Weese, Yasuyuki Sawada, and Yasutora Watanabe for their critical comments, and Shuhei Kainuma and Hiroto Sato for the continuous discussion and technical support. I am also grateful for useful comments and advice from Keisuke Kawata, Ayako Kondo, Chishio Furukawa, Daisuke Oyama, Drew Griffen, Rie Muraoka, and participants of the JADE young seminar. I also acknowledge PARLAP project members Oliver Coomes, Christian Abizaid, and Jennifer Langill. This study would not have been possible without the large-scale and thorough dataset collected by the PARLAP field team in Loreto (Carlos Rengifo Upiachihua, Judiht del Castillo Macedo, Norith Paredes Salas) and Ucayali (Luis Angel Collado Panduro, Claudio Sinuri Lomas, Santiago Nunta). Instituto Nacional de Innovación Agraria (INIA), Peru, and Bioversity International, Colombia thankfully provided the extra information for the systematic review.

†Graduate School of Economics, The University of Tokyo. y-noritomo@g.ecc.u-tokyo.ac.jp