Handling Limited Overlap in Observational Studies with Cardinality Matching[*]

Giancarlo Visconti[†]   José R. Zubizarreta[‡]

October 15, 2017

[†]Ph.D. Candidate, Department of Political Science, Columbia University, 420 West 118th Street, International Affairs Building, 7th Floor, New York, NY 10027 (email, giancarlo.visconti@columbia.edu).

[‡]Assistant Professor, Department of Health Care Policy, and Department of Statistics, Harvard University, 180 Longwood Avenue, Office 307-D, Boston, MA 02115 (email, zubizarreta@hcp.med.harvard.edu).

Abstract

A common problem encountered in observational studies is limited overlap in covariate distributions across treatment groups. To address this problem, and avoid strong modeling assumptions, it has become common practice to use matching methods and restrict the analyses to the portions of the treatment groups that overlap or, ultimately, that are balanced. Often, this is done by matching on the estimated propensity score or on coarsened versions of the observed covariates. A recent, alternative methodology that in a sense encompasses these two approaches is cardinality matching. Cardinality matching is a flexible matching method that uses integer programming to find the largest matched sample that is balanced as designed before matching by the investigator. In this paper, we apply and illustrate the method of cardinality matching and show how to use it to directly balance several features of the covariates, including their trajectories in time and their distributions, without requiring exact matching. We demonstrate how cardinality matching addresses the problem of limited overlap using the original covariates, as opposed to a summarized or coarsened version of them. We discuss how this method can be extended to build matched samples that are not only balanced but also representative by design. We also show how this method facilitates sensitivity analyses for hidden biases. We explain these advancements through an observational study of the electoral impact of the 2010 earthquake in Chile.

KEYWORDS: Causal inference; Matched sampling; Natural disasters; Natural experiments; Observational studies