

[DOI: 10.20472/IAC.2019.051.027](https://doi.org/10.20472/IAC.2019.051.027)

RAHUL MUKHERJEE

Indian Institute of Management Calcutta, India

TIRTHANKAR DASGUPTA

Rutgers University, United States

RANDOMIZATION-BASED CAUSAL INFERENCE FROM POSSIBLY UNBALANCED SPLIT-PLOT DESIGNS

Abstract:

Factorial experiments are currently undergoing a popularity surge in social and behavioral sciences. A key challenge here arises from randomization restrictions. Consider an experiment to assess the causal effects of two factors, expert review and teacher bonus scheme, on 40 schools in a state. A completely randomized assignment can disperse the schools undergoing review all over the state, thus entailing prohibitively high cost. A practical alternative is to divide these schools by geographic proximity into four groups called whole-plots, two of which are randomly assigned to expert review. The teacher bonus scheme is then applied to half of the schools chosen randomly within each whole-plot. This is an example of a classic split-plot design.

Randomization-based analysis, avoiding rigid linear model assumptions, is the most natural methodology to draw causal inference from finite population split-plot experiments as above. Recently, Zhao, Ding, Mukerjee and Dasgupta (2018, *Annals of Statistics*) investigated this for balanced split-plot designs, where whole-plots are of equal size. However, this can often pose practical difficulty in social sciences. Thus, if the 40 schools are spread over four counties with 8, 8, 12 and 12 schools, then each county is a natural whole-plot, the design is unbalanced, and the analysis in Zhao et al. (2018) is not applicable.

We investigate causal inference in split-plot designs that are possibly unbalanced, using the potential outcomes framework. We start with an unbiased estimator of a typical treatment contrast and first examine how far Zhao et al.'s (2018) approach can be adapted to our more general setup. It is seen that this approach, aided by a variable transformation, yields an expression for the sampling variance of the treatment contrast estimator but runs into difficulty in variance estimation. Specifically, as in the balanced case and elsewhere in causal inference (Mukerjee, Dasgupta and Rubin, 2018, *Journal of the American Statistical Association*), the resulting variance estimator is conservative, i.e., has a nonnegative bias. But, unlike most standard situations, the bias does not vanish even under strict additivity of treatment effects. To overcome this problem, a careful matrix analysis is employed leading to a new variance estimator which is also conservative, but enjoys the nice property of becoming unbiased under a condition even milder than strict additivity. We also discuss the issue of minimaxity with a view to controlling the bias in variance estimation, and explore the bias via simulations.

Keywords:

Bias, factorial experiment, finite population, minimaxity, potential outcome, variance estimation.

JEL Classification: C10, C13, C90