Department of Marketing

WU WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

Research Seminars Series

Eric Schwartz,
University of Michigan (US)

April 13th, 2023 | 13:00 - 15:00 | AD.0.090 Sitzungssaal 6

The A-B Test Deception: What Online Experiments Cannot (and Can) Tell You About How Customers Respond to Advertising

Experimenters, such as researchers and marketers, use tools provided by advertising platforms to conduct randomized experiments for testing user responses to creative elements in online ads. Internally valid comparisons between ads require the mix of experimental users exposed to each ad to be the same across all ads. But that internal validity is threatened when platforms' targeting algorithms deliver each ad to each ad's own optimized mix of users, where those mixes diverge across ads. We demonstrate how this method of targeting and delivering ads to users divergently across ads leads experimenters to incorrectly infer which ad performs better. We identify the conditions for this undetectable "Simpson's reversal," in which all unobserved types of users may prefer ad A over ad B, but the experimenter mistakenly infers from aggregate experimental results that users prefer ad B over ad A. due Since platforms aggregate test results across user characteristics and other factors used in targeting, this bias remains unobservable and uncorrectable by the experimenter without action from platforms. Through analysis and simulation, we quantify how much bias would decrease if platforms were to disable divergent delivery and instead use proportional targeting across ads, but also capture the loss of revenue generating activity for the platform. More generally, we characterize how bias in the aggregate estimate of the difference between two ads' lifts is driven by the interplay between heterogeneous responses to different ads and how platforms deliver ads to divergent subsets of users. Methodologically, we extend the potential outcomes model of causal inference to explicitly treat random assignment of ads and the user exposure states for each ad as two independent decisions.