

Decision Processes Colloquia

Monday, April 28, 2025

Where: JMHH G50

When: 12:00 – 1:20 pm

Joachim Vosgerau

Professor of Marketing
Bocconi

The Big Data Fallacy: When N is Large, Correlation = Causation

ABSTRACT:

The advent of big data has revolutionized how scientists, policy makers, and managers use empirical evidence to make decisions. Advocates of the big data revolution have propagated a dangerously misleading belief—that with enough data, the numbers speak for themselves. This claim overlooks a fundamental issue: big data are typically observational, rather than collected following the strict rules of random assignment. Consequently, causal relationships estimated using big data exhibit some degree of endogeneity, irrespective of sample size. Eight preregistered experiments (total N = 6,408) demonstrate what we term the "big data fallacy": decision-makers are more likely to erroneously infer causation from correlation in large-sample observational data. The fallacy persists across decision-domains and is exhibited even by those with moderately strong scientific reasoning skills and several years of decision-making experience. Two interventions to enhance the identification of endogeneity issues in large sample data failed to mitigate the big data fallacy.

