

Department SEMINARS



DEM DIPARTIMENTO DI
ECONOMIA E
MANAGEMENT

Inference for Regression with Variables Generated by AI or Machine Learning?

LAURA BATTAGLIA

University of Oxford

THURSDAY, 10 JULY 2025
2:00 PM

Online via Teams

A B S T R A C T

It has become common practice for researchers to use AI-powered information retrieval algorithms or other machine learning methods to estimate variables of economic interest, then use these estimates as covariates in a regression model. We show both theoretically and empirically that naively treating AI-and ML-generated variables as “data” leads to biased estimates and invalid inference. We propose two methods to correct bias and perform valid inference: (i) an explicit bias correction with bias-corrected confidence intervals, and (ii) joint maximum likelihood estimation of the regression model and the variables of interest. Through several applications, we demonstrate that the common approach generates substantial bias, while both corrections perform well.

For information: caterina.giannetti@unipi.it