



Achieving Reliable Causal Inference with Data-Mined Variables



Combining machine learning with econometric analysis is becoming increasingly prevalent in both research and practice. A common empirical strategy uses predictive modeling techniques to “mine” variables of interest from available data, then includes those variables into an econometric framework to estimate causal effects. However, because the predictions from machine learning models are inevitably imperfect, econometric analyses based on the predicted variables likely suffer from bias due to measurement error. We propose a novel approach to mitigate these biases, leveraging the ensemble learning algorithms to generate instrumental variables for bias correction. The random forest algorithm, for

example, performs best when comprised of a set of trees that are individually accurate in their predictions, yet which also make “different mistakes”, i.e., have weakly correlated prediction errors. A key observation is that these properties are closely related to the relevance and exclusion requirements of valid instrumental variables.

Edward McFowland III, Ph.D.

Assistant Professor of Business
Administration
Harvard Business School

Date: Monday, March 06, 2023

Time: 2:00 – 3:00 p.m. US Central Time

Location: **Wehner (WCBA) 360K** (in person)

Online: 998 4499 3279 (ID) & 724615 (PWD)

Faculty host: Bin Zhang, INFO

Biography

Dr. Edward McFowland III is an Assistant Professor in the Technology and Operations Management Unit at Harvard Business School. Dr. McFowland’s research interests—which lie at the intersection of Machine Learning, Information Systems, and Management—include the development of computationally efficient algorithms for large-scale statistical machine learning and “big data” analytics. As a data and computational social scientist, Dr. McFowland aims to bridge the gap between machine learning and the social sciences (e.g., economics, public policy, and management). His work has been published in leading management, machine learning, and statistics journals, and has been supported by Adobe, Facebook, PNC Bank, AT&T Labs, and NSF. Dr. McFowland earned his Ph.D. in Information Systems and Management from Carnegie Mellon University. He also holds Masters degrees in Machine Learning, Public Policy, and in Information Systems from Carnegie Mellon University. Prior to joining HBS, Dr. McFowland taught at the University of Minnesota Carlson School of Management.

You can also click this link to join the seminar <https://tamu.zoom.us/j/99844993279?pwd=TKlodWFVRURyMmMkwakl4SWZGeVlTQT09>

