

## Machine Learning for Causal Inference

Tyler Lenk, Class of 2026

During my abroad semester at the University of Edinburgh in Fall 2024, I took two core empirical economics courses: Econometrics and Data Science. Both rely on regression — the canonical technique for modeling and estimating relationships between variables — but they answer distinct questions with different approaches. While Econometrics focuses on credibly estimating how an intervention may change an outcome (i.e. causal inference), Data Science aims predict outcomes given input data without regard for underlying mechanisms (i.e. machine learning (ML)). Due to this disconnect, many economists are skeptical of the application of ML to econometrics. However, in my last week in Edinburgh, I learned of new ML methods designed to aid econometrics. Excited by the new methods, I made it the goal of my summer to understand and apply them to a prior economics paper and report on their effectiveness.

To evaluate a new ML econometric method, I first had to contextualize its place in the history of the field. I parsed review articles and seminal econometrics papers to create a timeline of key developments in causal inference. Then, I searched the causal ML literature and identified the most prominent new developments. For my empirical project, I chose to implement a strategy for inference for treatment effects after ML-driven variable selection (known as Post-Double-Selection Lasso, or PDS Lasso for short).<sup>1</sup>

What is the PDS Lasso used for? Econometricians often wish to estimate treatment effects in observational contexts — but, due to the lack of randomization, they typically must control for variables “confounded” with the causal effect of the treatment on the outcome. In situations with many potential control variables (or confounders) and limited data points, only a few controls can be included in the model. PDS Lasso uses machine learning to select the most important controls and provides valid measures of estimate uncertainty following the procedure.

To test the PDS Lasso, I chose a paper that estimates the effect of market competition on productivity for developing countries.<sup>2</sup> Economic theory informs us that firms in markets with more competitors are pushed harder to innovate. However, estimation of this all-else-constant effect of competition on productivity requires careful econometrics, including selection of controls. The original paper used publicly available firm-level survey data from the 2004 World Bank Enterprise Survey. I gained access to this survey data, and I am currently conducting a re-analysis using STATA and R. The project is ongoing, and I have plans to implement ML methods I learned this summer for my honors thesis as well.

**Faculty Mentor: Mahdi Kahou**

**Funded by the Surdna Foundation Undergraduate Research Fellowship Program**

---

<sup>1</sup> Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81(2), 608-650.

<sup>2</sup> Schiffbauer, M., & Ospina, S. (2010). *Competition and firm productivity: Evidence from firm-level data* (No. 2010/067). International Monetary Fund.