Update to Analysis of College Athletic Success using Machine Learning

Daniel Pate, Gerard Tetegan, Jacob Lara, and Lane Whitehead
Dr. M. Jahangir Alam, Department of Economics, Texas A&M University, Spring 2024

Abstract

This study improves the causal estimation workflow in Anderson (2017) by first re-estimating the propensity scores using Artificial Neural Networks (ANN) and then integrating Ridge regression and bootstrapping techniques to address multicollinearity among the predictors. This ensures robust and comparable conclusions to the Anderson (2017) research, enhancing the reliability and consistency of the findings.

Introduction

- Anderson (2017) employed propensity score weighting to address endogeneity by estimating scores for each observation and using them as weights in regression.
- This extension employs ANN for propensity score estimation, and Ridge regression with bootstrapping for causal inference.

Literature Review

- ANN promising for propensity score estimation because they algorithmically handle nonlinear relationships and interactions (Keller et al. 2015).
- Ridge regression is recommended for scenarios with fewer predictors where each predictor is expected to significantly influence predictions (Xu, W., 2019. Towards Data Science).
- Bootstrap techniques enhance the statistical inference capabilities of ridge regression models by providing more accurate and reliable standard error estimations, particularly in the presence of highly correlated predictors (Capur, 2023)

Methodology

- ANN model used to capture non-linearity in the data the logistic regression tried to incorporate.
- Weighted Ridge Regression was used to addressing multicollinearity effectively and endogeneity.
- Bootstrapping was used to estimate standard errors and confidence intervals.

Replication Results

The tables below show the comparison between Table 3 from Anderson (2017) alongside the results of the replication in **Python**.

Anderson (2017)

Outcome	Coefficient	SE	N
Alumni Athletic Operating Donations	191.2	65.0	616
Alumni Nonathletic Operating Donations	-137.4	96.1	616
Total Alumni Donations	267.4	267.1	1258
Alumni Giving Rate	0.0002	0.0007	1287
Academic Reputation	0.003	0.002	650
Applicants	81.1	60.4	528
Acceptance Rate	-0.003	0.002	979
First-Time Out-of-State Enrollment	1.6	5.0	962
First-Time In-State Enrollment	12.6	6.4	962
25th Percentile SAT	0.8	0.7	426

Python Replication

Outcome	Coefficient	SE	N
Alumni Athletic Operating Donations	191.2	65.0	616
Alumni Nonathletic Operating Donations	-137.4	96.1	616
Total Alumni Donations	267.4	267.1	1258
Alumni Giving Rate	0.0002	0.0007	1287
Academic Reputation	0.003	0.002	650
Applicants	81.1	60.4	528
Acceptance Rate	-0.003	0.002	979
First-Time Out-of-State Enrollment	1.6	5.0	962
First-Time In-State Enrollment	12.6	6.4	962
25th Percentile SAT	0.8	0.7	426

PSM Grouping (Logit vs ANN)

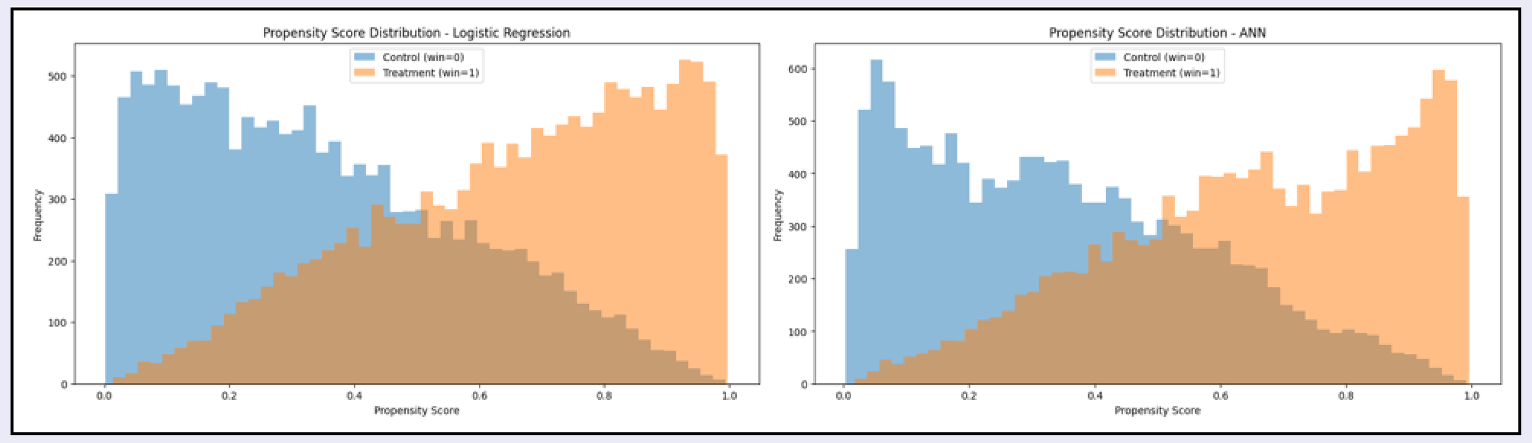

Figure: PSM Grouping (Logit vs ANN)

Machine Learning

Causal Estimation using Ridge Regression.

Outcome	Coeff	SE	Lower CI	Upper CI
Alumni Athletic Op	107.8	31.3	50.8	175.1
Alumni Nonathletic Op	44.9	54.0	-62.7	151.5
Total Alumni Donations	143.5	144.1	-146.1	418.2
Alumni Giving Rate*	0.4	0.4	-0.4	1.1
Academic Rep*	1.1	1.2	-1.3	3.3
Applicants	39.6	36.3	-32.2	110.4
Acceptance Rate	-0.3	0.1	-0.5	-0.1
First Time (OoS)	2.4	2.3	-2.1	6.7
First Time (IS)	15.6	4.4	7.1	24.3
SAT	0.4	0.4	-0.4	1.1

*multiplied by 1000 for clarity.

Conclusion

- The coefficient for donations to Athletics and the first-time in-state enrollment show significant changes of -43% and 24% respectively.
- More wins lead to lower acceptance rates.

References

- Anderson, M. (2017). The Benefits of College Athletic Success: An Application of the Propensity Score Design with Instrumental Variables. Review of Economics and Statistics, 99. DOI: 10.1162/REST_a_00589.
- Keller, B., Kim, J.-S., & Steiner, P. (2015). Neural Networks for Propensity Score Estimation: Simulation Results and Recommendations. In Handbook of Statistics: Data Science, Time Series and Dynamic Models (pp. 279-291). Springer. DOI: 10.1007/978-3-319-19977-1_20.

Appendix

Figure: GitHub;References