Update to Analysis of College Athletic Success using Ridge Regression

Daniel Pate, Gerard Tetegan, Jacob Lara, and Lane Whitehead
Dr. M. Jahangir Alam, ECMT 680 - Financial Econometrics, Texas A&M University, Spring Semester 2024

Abstract

This study improves the causal estimation workflow in the original paper by integrating Ridge regression and bootstrapping techniques to address multicollinearity among predictors, a common challenge in traditional regression analyses. The integration of these methods ensures robust and comparable conclusions to the original research, enhancing the reliability and consistency of the findings.

Introduction

- Original paper employed propensity score weighting to address endogeneity by estimating scores for each observation and using them as weights in regression.
- This extension employs ridge regression with bootstrapping for causal inference, while still incorporating propensity scores to address endogeneity concerns.

Literature Review

- 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

- Replication utilized weighted linear regression in Python, mirroring the methodology employed in STATA.
- Propensity scores were estimated through logistic regression.
- For the machine learning implementation, Ridge Regression was applied to replicate a straightforward and interpretable model, addressing multicollinearity effectively.
- Bootstrapping was employed to estimate standard errors, facilitating direct comparison with results obtained using weighted least squares (WLS).

Replication Results

The tables below show the comparison between Table 3 (from the original paper) alongside the results of the replication in Python.

TABLE 3.—EFFECTS OF FOOTBALL WINS ON OUTCOMES

	STE Model	
Outcome	Coefficient	*N*
Alumni Athletic Operating Donations	191.2 (65.0)	616
Alumni Nonathletic Operating Donations	−137.4 (96.1)	616
Total Alumni Donations	267.4 (266.9)	1,258
Alumni Giving Rate	0.0002 (0.0007)	1,287
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

Figure: Original

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

Figure: Replication

Machine Learning

Causal Estimation using Ridge Regression.

Outcome	Coefficient	SE	N
Alumni Athletic Operating Donations	190.5	34.6	616
Alumni Nonathletic Operating Donations	-138.2	117.6	616
Total Alumni Donations	269.8	215.1	1258
Alumni Giving Rate	0.0002	0.0009	1287
Academic Reputation	0.003	0.002	650
Applicants	80.4	121.0	528
Acceptance Rate	-0.003	0.002	979
First-Time Out-of-State Enrollment	1.6	7.2	962
First-Time In-State Enrollment	12.7	10.8	962
25th Percentile SAT	0.8	1.7	426

Figure: Ridge Regression

Conclusions

- Both Ridge and OLS deliver closely aligned coefficients, affirming its effectiveness in traditional analysis contexts.
- Consistency in coefficient magnitude across methods underlines the robust control of multicollinearity, enhancing the reliability of the statistical findings.

References

- Anderson, M. (2012). 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.
- Özkale, M. R., & Altuner, H. (2023). Bootstrap confidence interval of ridge regression in linear regression model: A comparative study via a simulation study. Communications in Statistics - Theory and Methods, 52(20), 7405-7441. DOI: 10.1080/03610926.2022.2045024.

Appendix

SCAN ME

Figure: GitHub;References