

Revisiting PSM Analysis of College Athletic Success with Machine Learning: LASSO Regression & Gradient Boosting

Gerard Tetegan, Jacob Lara, Daniel Pate, and Lane Whitehead
Texas A&M University
April 29, 2024

Abstract

This study PSM to determine the causal effects of college football success on donations and applications. We replicated their analysis and then used LASSO regression & (possibly) gradient boosting aiming to improve the robustness of the propensity scores.

Introduction

The study uses bookmaker spreads to calculate the propensity scores and found that football success leads to better outcomes. We use LASSO and gradient boosting to aiming improve the PSM estimates and variable selection to see their effects.

Literature Review

- Pairing PSM and gradient boosting significantly enhances predictive accuracy (Kim et al., 2023)
- Gradient boosting is more robust and preferred over other ML methods in finding average treatment effects (Yang, Chuang, Kuan, 2020)
- LASSO regressions have demonstrated their usefulness in subset selection especially in high-dimensional settings (Tibshirani, 1996).
- While LASSO is not as powerful in comparison to other ML methods, combined with PSM it demonstrates extreme robustness (Pirracchio, Petersen, van der Laan, 2015)

Findings

| | coef | std err | z | P> z | [0.025 | 0.975] |
|--|---------|---------|--------|-------|---------|--------|
| lead2_pscore_wk11_group_1 | 0.1529 | 0.048 | 3.182 | 0.001 | 0.059 | 0.247 |
| lead2_pscore_wk11_group_2 | -0.0803 | 0.309 | -0.260 | 0.795 | -0.686 | 0.525 |
| lead2_pscore_wk11_group_3 | 1.6379 | 1.016 | 1.613 | 0.107 | -0.353 | 3.628 |
| lead2_pscore_wk11_group_4 | 0.6030 | 0.810 | 0.744 | 0.457 | -0.985 | 2.191 |
| lead2_pscore_wk11_group_5 | 0.5882 | 2.390 | 0.246 | 0.806 | -4.097 | 5.273 |
| lead2_pscore_wk11_group_6 | 2.1827 | 0.586 | 3.722 | 0.000 | 1.033 | 3.332 |
| lead2_pscore_wk11_group_7 | 0.6534 | 1.576 | 0.415 | 0.678 | -2.435 | 3.742 |
| lead2_pscore_wk11_group_8 | 0.0269 | 1.503 | 0.018 | 0.986 | -2.920 | 2.974 |
| lead2_pscore_wk11_group_9 | -4.1408 | 2.235 | -1.853 | 0.064 | -8.520 | 0.239 |
| lead2_pscore_wk11_group_10 | 1.9579 | 1.815 | 1.078 | 0.281 | -1.600 | 5.516 |
| lead2_pscore_wk11_group_11 | -0.4022 | 1.408 | -0.286 | 0.775 | -3.162 | 2.358 |
| lead2_pscore_wk11_group_12 | -0.6552 | 0.701 | -0.935 | 0.350 | -2.029 | 0.719 |
| lead2_pscore_wk11_group_1*lead2_win_wk11 | -0.0292 | 0.136 | -0.215 | 0.830 | -0.295 | 0.237 |
| lead2_pscore_wk11_group_2*lead2_win_wk11 | 0.0637 | 0.139 | 0.459 | 0.646 | -0.208 | 0.336 |
| lead2_pscore_wk11_group_3*lead2_win_wk11 | -0.1657 | 0.196 | -0.845 | 0.398 | -0.550 | 0.218 |
| lead2_pscore_wk11_group_4*lead2_win_wk11 | -0.0570 | 0.080 | -0.710 | 0.478 | -0.214 | 0.100 |
| lead2_pscore_wk11_group_5*lead2_win_wk11 | -0.0352 | 0.186 | -0.189 | 0.850 | -0.400 | 0.330 |
| lead2_pscore_wk11_group_6*lead2_win_wk11 | 0.0425 | 0.066 | 0.643 | 0.521 | -0.087 | 0.172 |
| lead2_pscore_wk11_group_7*lead2_win_wk11 | 0.2803 | 0.138 | 2.029 | 0.043 | 0.009 | 0.551 |
| lead2_pscore_wk11_group_8*lead2_win_wk11 | 0.0587 | 0.091 | 0.647 | 0.518 | -0.119 | 0.237 |
| lead2_pscore_wk11_group_9*lead2_win_wk11 | 0.0580 | 0.091 | 0.636 | 0.525 | -0.121 | 0.237 |
| lead2_pscore_wk11_group_10*lead2_win_wk11 | -0.0098 | 0.105 | -0.093 | 0.926 | -0.215 | 0.196 |
| lead2_pscore_wk11_group_11*lead2_win_wk11 | 0.0139 | 0.092 | 0.151 | 0.880 | -0.167 | 0.195 |
| lead2_pscore_wk11_group_12*lead2_win_wk11 | 0.0805 | 0.117 | 0.688 | 0.491 | -0.149 | 0.310 |
| lead2_pscore_wk11_group_1*lead2_pscore_wk11 | 2.5346 | 0.773 | 3.280 | 0.001 | 1.020 | 4.049 |
| lead2_pscore_wk11_group_2*lead2_pscore_wk11 | 2.5086 | 1.807 | 1.388 | 0.165 | -1.033 | 6.050 |
| lead2_pscore_wk11_group_3*lead2_pscore_wk11 | -4.4717 | 3.862 | -1.158 | 0.247 | -12.041 | 3.098 |
| lead2_pscore_wk11_group_4*lead2_pscore_wk11 | -0.3736 | 2.599 | -0.144 | 0.886 | -5.468 | 4.721 |
| lead2_pscore_wk11_group_5*lead2_pscore_wk11 | -0.2291 | 6.416 | -0.036 | 0.972 | -12.804 | 12.346 |
| lead2_pscore_wk11_group_6*lead2_pscore_wk11 | -3.9546 | 1.326 | -2.981 | 0.003 | -6.554 | -1.355 |
| lead2_pscore_wk11_group_7*lead2_pscore_wk11 | -0.4800 | 2.823 | -0.170 | 0.865 | -6.013 | 5.053 |
| lead2_pscore_wk11_group_8*lead2_pscore_wk11 | 0.7010 | 2.536 | 0.276 | 0.782 | -4.270 | 5.672 |
| lead2_pscore_wk11_group_9*lead2_pscore_wk11 | 6.8924 | 3.349 | 2.058 | 0.040 | 0.329 | 13.456 |
| lead2_pscore_wk11_group_10*lead2_pscore_wk11 | -2.0122 | 2.446 | -0.823 | 0.411 | -6.806 | 2.781 |
| lead2_pscore_wk11_group_11*lead2_pscore_wk11 | 1.1478 | 1.735 | 0.662 | 0.508 | -2.253 | 4.548 |
| lead2_pscore_wk11_group_12*lead2_pscore_wk11 | 1.2540 | 0.796 | 1.575 | 0.115 | -0.307 | 2.814 |

Figure: Simple OLS Regression

| | coef | std err | z | P> z | [0.025 | 0.975] |
|---|---------|---------|--------|-------|--------|--------|
| const | 0.6200 | 0.030 | 20.782 | 0.000 | 0.561 | 0.678 |
| lead2_pscore_wk10_group_1 | -0.2114 | 0.037 | -5.694 | 0.000 | -0.284 | -0.139 |
| lead2_pscore_wk10_group_2 | -0.0790 | 0.046 | -1.723 | 0.085 | -0.169 | 0.011 |
| lead2_pscore_wk10_group_3 | -0.0775 | 0.053 | -1.461 | 0.144 | -0.181 | 0.026 |
| lead2_pscore_wk10_group_5 | 0.0691 | 0.050 | 1.393 | 0.163 | -0.028 | 0.166 |
| lead2_pscore_wk10_group_9 | 0.0441 | 0.094 | 0.471 | 0.638 | -0.139 | 0.228 |
| lead2_pscore_wk10_group_10 | 0.0313 | 0.075 | 0.416 | 0.677 | -0.116 | 0.179 |
| lead2_pscore_wk10_group_12 | 0.1960 | 0.099 | 1.987 | 0.047 | 0.003 | 0.389 |
| lead2_pscore_wk10_group_8*lead2_win_wk10 | 0.1996 | 0.074 | 2.697 | 0.007 | 0.055 | 0.345 |
| lead2_pscore_wk10_group_9*lead2_win_wk10 | 0.0849 | 0.111 | 0.768 | 0.443 | -0.132 | 0.302 |
| lead2_pscore_wk10_group_10*lead2_win_wk10 | 0.1574 | 0.085 | 1.858 | 0.063 | -0.009 | 0.323 |
| lead2_pscore_wk10_group_11*lead2_win_wk10 | 0.1579 | 0.056 | 2.821 | 0.005 | 0.048 | 0.268 |
| lead2_pscore_wk10_group_12*lead2_win_wk10 | 0.0663 | 0.099 | 0.667 | 0.505 | -0.129 | 0.261 |

Figure: LASSO Regression

Machine Learning

Total observations in sample: 396
Mean Coefficients Comparison (OLS vs Lasso):
Week 1: OLS Mean = 3.159864011551365, Lasso Mean = 0.7373896301618871
Week 2: OLS Mean = 2.5512456523963376, Lasso Mean = 4.148783250009387
Week 3: OLS Mean = 1.6978220772375572, Lasso Mean = 0.3889874240731325
Week 4: OLS Mean = 1.7704137261095378, Lasso Mean = 1.062305113324602
Week 5: OLS Mean = 2.007464100828, Lasso Mean = 1.1030511866151986
Week 6: OLS Mean = 1.0554369496122213, Lasso Mean = 0.4642468134651469
Week 7: OLS Mean = 1.0564392198599402, Lasso Mean = 0.5368803872698791
Week 8: OLS Mean = 0.813826688044924, Lasso Mean = 0.14782238695137892
Week 9: OLS Mean = 0.6034733774968554, Lasso Mean = 0.3768318950707437
Week 10: OLS Mean = 0.09485391699786991, Lasso Mean = 0.0968232556146992
Week 11: OLS Mean = 0.1761845245210503, Lasso Mean = 0.10096149894181808
Week 12: OLS Mean = 0.0, Lasso Mean = 0.0
Paired t-test results:
T-statistic: 1.7856785382855855
P-value: 0.10171519886840388
There is no statistically significant difference in mean coefficients (fail to reject null hypothesis).

Figure: T-Test for Coefficients

Conclusions

The study finds that success in college football positively affects alumni donations, application rates, and academic reputation, while also increasing in-state enrollment and SAT scores of incoming students, especially in elite conferences. Our machine learning technique shows that with LASSO while some variables are omitted in the earlier stages it does not change the factors that are overall chosen.

Appendix



Figure: Appendix/Github