

# 1 Project Outline

- Introduction/Background

- Begin by discussing the mammoth investments that many colleges make into their football programs despite the fact that most athletic departments fail to pull in revenue that exceeds operating costs each year. Address the arguments on both sides of this issue. Should schools really be putting so many resources towards something that doesn't directly support their primary mission to educate? What about the supposed indirect benefits that would certainly contribute to this mission if they are true?
- Should the football program perform well, these hypothesized ancillary benefits include the following: more applicants, more enrollments (and higher yield), higher quality students (e.g. in terms of SAT scores), higher freshman retention and graduation rates, and more funding via increased alumni donations and/or state appropriations. This paper looks to test the applicant, admit, enrollment, and yield hypotheses in particular, though some room may be given to explain why some people think each of these outcomes might occur. Discuss the theoretical underpinnings; Ex: the (ideally successful) athletics program as a university's "front porch" theory, as a signal for overall university quality (advertising effect), as an amenity like nice residence halls and libraries, as a means to socially integrate the student body and build a sense of camaraderie, as a "stock of goodwill", as a means for students to relieve stress as one would at other party-esque social events, as a means to increase student body diversity, etc. The general idea is that people have offered a number of reasons why college football programs (successful ones in particular) may actually help improve university academics.
- Mention anecdotal evidence of football and men's basketball performance seemingly impacting admissions (e.g. after Appalachian State vs. Michigan, after Oklahoma vs Boise State, etc.); pertains particularly to cases in which massive underdogs won a big game or a traditionally weak program rose up to the top rungs (e.g. TCU).
- This paper will look to focus on Power Five conference teams only since we might reasonably hypothesize that these would be the most likely to reap the aforementioned indirect benefits of college football success (should they exist) due to their visibility (they get the most coverage from TV outlets and other sports media), name-recognition, and collective potential for success (for example, teams outside of this group have virtually no shot at winning a national title). The dependent variables of interest will be number of applications, quality of incoming freshman class (via SAT scores), acceptance rate, and yield. A fixed effects model is used primarily since it allows us to control for school-specific characteristics that we

might consider time-invariant over the span of our data. We use two separate panel data sets for the 65 Power Five teams: One for the BCS era and one for the present Playoff era; it may be interesting to see if there is any difference on the performance variables' coefficients between eras (after major conference realignment, new TV contracts, and the switch to a different post-season system). More information on the decision to look at this set of years in particular is given in the literature review.

- Lit Review

- Focus particularly on empirical results from the literature that explores the link between college athletic program (football or men's basketball) performance and at least one of the following variables: applicant and admit pool size, incoming freshman quality (in terms of SAT scores), yield, and acceptance rate.
- Introduce some of the seminal papers in chronological order.
- Address the apparent controversy surrounding what can be taken from the previous literature as of now. Two very recent papers Eggers et al. (2020) and Baumer and Zimbalist (2019) disagree considerably; the former states that the literature “overwhelmingly suggests that athletic success positively influences both the quantity and quality of students at a university” while the latter states that it gives “some support to the notion that robust athletic success can lead to an increase in applications to a school”, but shows “only weak support, if any, for the claim that sport success leads to an increase in the quality of students”; Baumer and Zimbalist (2019) point out what they believe to be problems in the empirical analysis methodology and statistical inferences made in some papers. Furthermore, note that Pope and Pope (2009) also mention that the literature in this area has produced “inconsistent results” and explain that this may be due to differences in empirical model choices (e.g. cross-sectional OLS vs. fixed effects regression) and model specification choices (e.g. what variables should be used to measure athletic success and what control variables should be employed). We might also add to this that many of the studies use different data sets (e.g. different spans of time and different sets of teams).
- Baumer and Zimbalist (2019) point out that almost all of the previous studies looked at data from before the 2000s. Hence, this paper may be useful from the standpoint that it considers two sets of data over 1998-2013 and 2014-2020. This more recent data set captures many important changes to the college football landscape (e.g. more significant TV contracts—resulting in increased visibility and revenue—and the corresponding heightened emphasis on athletic program development and management among Power Five conference universities), some of which were mentioned above.

- Data

- Integrated Postsecondary Education Data System (IPEDs): acceptance rate, yield, applicant pool size, SAT scores for incoming freshman.

- National Center for Education statistics (NCES) and Census Bureau: number of high school graduates by state, median per capita or household income by state.
- Could potentially include more control variables that may influence the decision to apply/enroll (particularly those that relate to overall school quality). Pope and Pope (2009) includes two additional variables that are thought to control for differences in school quality: average full-time professor salary and total annual cost of attendance; they mention robustness testing with numerous other school quality controls (e.g. number of national merit scholars, student-faculty ratio, percentage of students that go to graduate school, etc.) and find that their results are essentially the same. They also note that using these additional controls results in major statistical power losses due to missing values. One thing to note too is that Pope and Pope (2009) were also using a fixed effects model that controlled for time-invariant school-specific characteristics as well as nationwide trends through a year dummy. It might actually be reasonable to say that the big components of school quality that may influence applicant decisions (e.g. general perception of academic and/or social reputation) are more or less unchanging—especially if the time span your panel data set covers isn’t very long. Consider that Baumer and Zimbalist (2019) do not include these types of controls in their fixed effects models; they note that they tested them with cost of attendance and other variables, but ran into issues with missing data and high multicollinearity. Cost of attendance could also inform decisions simply based upon ability to pay, but we are already electing to control (at least weakly) for differences in income; also, tuition costs have likely been rising in a similar way for all of the colleges in our sample over the past few decades.

- Empirical Methods

- Fixed effects regression models. Primarily to control for unobserved characteristics that vary by school (e.g. team name-recognition, cultural traditions surrounding athletics and/or academics, perception of school’s academic reputation and/or social scene, etc.) that may be seen as essentially invariant across the years of our panels. This also holds the added benefit of controlling for observable characteristics that are certainly time-invariant for the purpose of our data set (e.g. whether the school is public, private, or religiously-affiliated, geographic location, etc.). We also add in a year dummy to control for macroeconomic trends that may affect the response variables (e.g. inflation, GDP changes, etc.).
- Different dependent variables to consider: number of applicants, quality of enrollees (by SAT scores), acceptance rate, and yield.
- Explanatory variables (indicators of college football success): overall win-loss ratio, final AP poll top 15 dummy, final AP poll top 5 dummy, national title winner dummy (through BCS or playoff system depending on the data set). We use two lags on these success indicator variables to account for the timing of when students typically apply to colleges and when the college football season ends; lags also allow us to look at the residual impact of success on the response variables

(e.g. we can see if (and how) the national title win three years ago affects the response variables).

- Non-fixed control variables: number of high school graduates in a given state (rough measure of size of pool of potential applicants in a given year, especially for public colleges in the sample) and median per capita income (rough measure of average ability to pay for college; may also be more likely to apply to college if this is higher overall).
  - Note model assumptions.
  - May include Hausman test to see if a random effects model is more appropriate (though we should imagine that the unobserved time-invariant school effect IS correlated with our X's—especially the success indicator variables).
  - Display VIF to see if there are any multicollinearity problems.
  - Use standard errors that are robust to heteroskedasticity and serial correlation within a given school; clustered standard errors. According to Baumer and Zimbalist (2019), we may also need to adjust the p-values for multiple tests.
- (Anticipated) findings
    - They may be similar to those in Baumer and Zimbalist (2019) since we both consider Power Five teams only over a similar time frame and our regression models are similar (the main difference is the segmentation of the data set based upon BCS or playoff era status and the type of success indicator variables we use).
    - Note that the model is not capable of considering the impact of sustained college football success on the response variables (e.g. a team that is consistently in the top 5 each year); it only looks at the average impact of changes in success across seasons. If the success variables were not to change from year to year for a given team, then there would be no effect to observe (remember that the time-invariant school term is purposely removed by construction of the fixed effects model). This might be accomplished via cross-sectional OLS, but the issue of too few data points may arise.
  - Conclusion
    - TBD depending on findings.