Winning in the Margins: Using Marginal Revenue and Pythagorean Wins to Value College Quarterbacks

Joey Larabee

Introduction

In 2019 the state of California passed legislation that allowed college athletes playing for NCAA programs to profit off of their Name, Image, and Likeness. This was the first domino to fall leading to today's environment where schools are now required to share revenue with their athletes, and donor-sponsored NIL deals are exploding in popularity. The current environment is commonly referred to as the "Wild West" as players, agents, schools, and donors seem to be acting with no rationale. This is the type of open market for talent is a ripe environment to be capitalized on for those who can spend their money most efficiently. To behave more efficiently I would propose a system for valuing player talent, and specifically quarterbacks, by connecting their contributions on the field to the dollars collected off the field.

In 1985 Bill James introduced his Pythagorean Win Formula for baseball, theorizing that there is a proportional relationship between the runs a team gave up and scored, and the amount of games that they won. Across a few sports now statisticians have attempted to apply this formula, solve for the different variables, and predict their own seasons. For the purposes of this project, equation 1 will represent the pythagorean wins formula.

$$\text{Win\%} = \frac{\text{Points For}^{\gamma}}{\text{Points For}^{\gamma} + \text{Points Against}^{\gamma}} \tag{1}$$

With Points For, Points Against, and Win Percentage all being variables that can be observed in a season, this only leaves Gamma, or the implied exponent as the only unobserved variable. However, solving for this implied exponent will only give us the relationship between points and winning. To take us all the way to a dollar valuation of the contribution of a single player we will need to then isolate any change in points to a projected change in revenue for the team. To do this I propose to use Duane Rockerbie's formula that he used in 2010 to estimate baseball player contracts, but rather than his logistic formula for player contributions, I propose the use of the Pythagorean implied exponent. The Marginal Revenue created by a player would look like this:

$$\text{MRP} = MR\gamma Y(1-Y)\left(\frac{dPF}{PF} - \frac{dPA}{PA}\right) \tag{2}$$

Where MR is the marginal revenue gained by the school for one extra win, Y is the team's Pythagorean winning percentage, and finally the impact of a single player on a team's points for and against. The implied exponent and a team's Pythagorean winning percentage are both calculated above, but for this project the marginal revenue of a singular win will also need to be calculated.

Data

For this project I sourced data from three distinct sources. For the college football performance data I utilized the 'cfbfastr' package in R to source data from collegefootballdata.com's publicly available API. To explore collegiate athletics revenue data I sourced information from the Knight-Newhouse College Athletics Database, which is a collaborative effort between the Knight Commission on Intercollegiate Athletics and Syracuse University's Newhouse School to provide clear and transparent financial data on collegiate athletics. Finally, enrollment data was sourced from the National Center for Education Statistics Integrated Postsecondary Education Data System. After joining our datasets together the first thing I examined was the correlation's between our features to examine if the hypothesis that revenue has a direct correlation to winning football games.

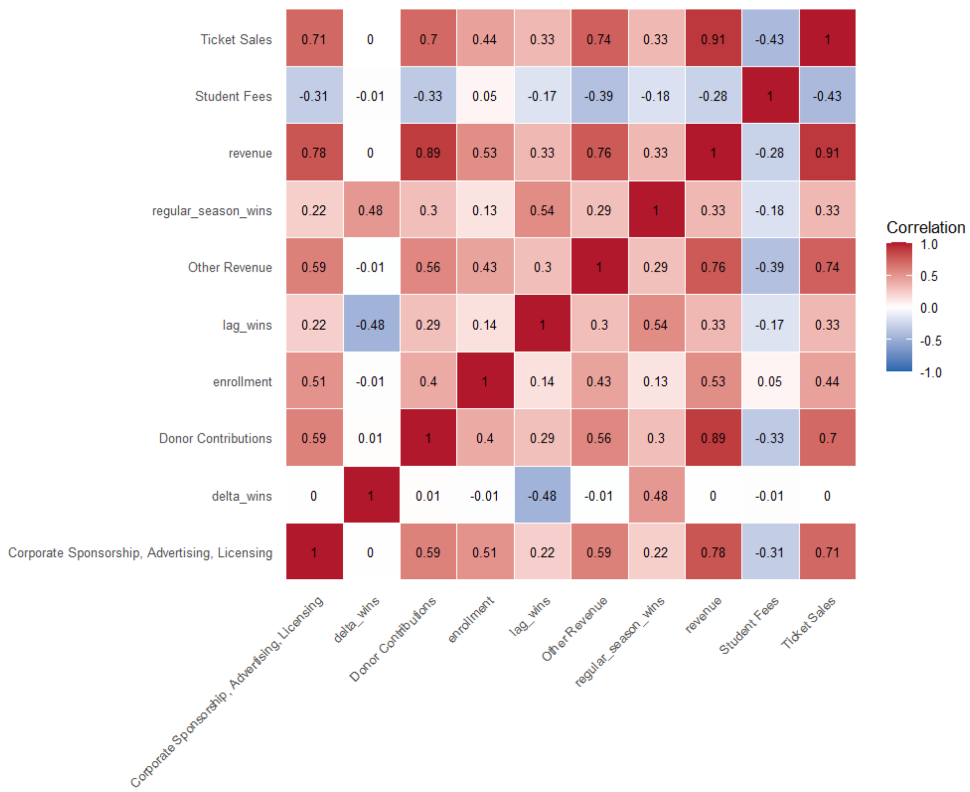

Figure 1: Correlation Matrix

In figure 1 we see a correlation matrix between our important features for modeling revenue as a function of wins from every school in the dataset. The first important takeaway we can draw from this is that there is a loose correlation between regular season wins and school revenue, especially our largest drivers of athletics revenue, ticket sales and donor contributions. However we do see a stronger correlation to revenue from overall school enrollment, meaning that if we wish to isolate the impact of football wins on athletic revenue we will have to adjust for a school's size. These are, however, encouraging results as football is just one sport at schools that often have at least 10, and these schools operate in such vastly different economic environments.

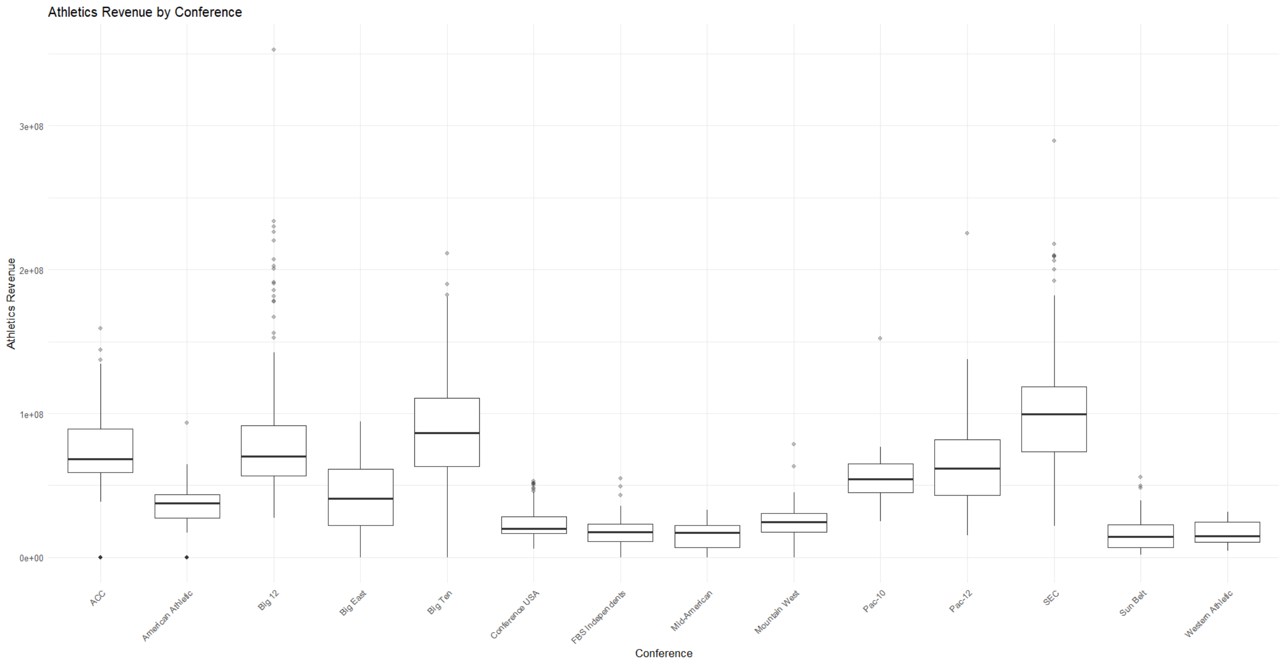

Figure 2: Revenue by Conference

Speaking of different economic environments, it will be important to note that the conference a team plays in will have a significant impact on their earnings potential. College football is commonly split in to 2 different categories, the Power 4 and the Group of 5. The Power 4 schools are the schools that play in the four most powerful conferences, SEC, Big Ten, Big 12, and ACC, while the Group of 5 is used to refer to the remaining schools. In figure 2 what we notice is that across the Group of 5 schools (G5) their revenue is rather similarly distributed, while the Power 4 schools (P4) schools each tend to have their own unique distributions. Due to these differences this project will focus on modeling specifically P4 team revenues using their conferences as a factor in the model.

Methods

To begin the framework of this equation we must first solve for the implied gamma from each season of team performance data we have available. Equation 3 below is the original Pythagoren wins equation applied to each season of college football going back to 2005.

$$\gamma_{\text{implied}} = \frac{\log\left(\dfrac{1 - \text{Win\%}}{\text{Win\%}}\right)}{\log\left(\dfrac{\text{Points Against}}{\text{Points For}}\right)} \tag{3}$$

After applying this method to all our seasons we see the distribution of all implied exponents collect rather distinctly to a centralized mean, our first promising indication that the Pythagorean method will fit for predicting college football seasons.

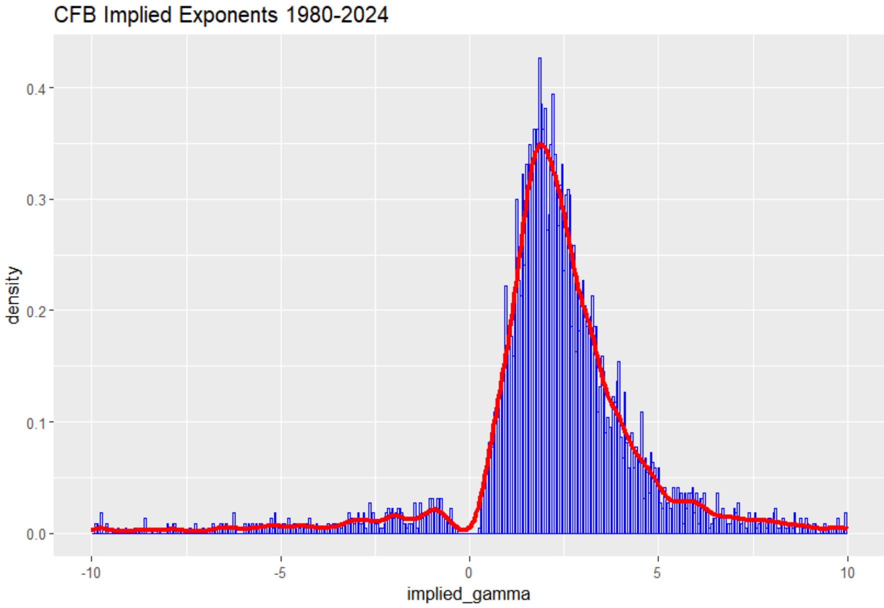

Figure 3: Distribution of Implied Exponents

For the purposes of this project we will use the mean of this distribution (2.559) in place of Gamma, and as a sanity check when we use this as gamma to predict the winning percentages of our observed seasons we get a correlation coefficient of 0.92 and a distinctly linear relationship.

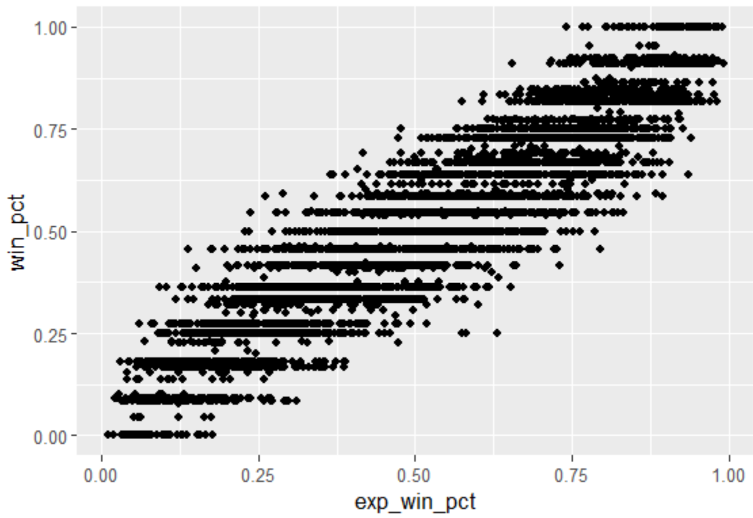

Figure 4: Pythag Win Pct vs Observed Win Pct

Looking back at equation 2 we now have our implied γ, with which we can then calculate our Y, and then we will use the Predicted Points Added value calculated by collegefootballdata.com as a stand in for a Quarterback's individual impact on Points For. This leaves us with our final variable to estimate: Marginal Revenue.

To calculate the marginal revenue added for each win I decided to utilize a linear regression model based on predicting total athletic revenue as a function of regular season wins, total enrollment, and what conference the team plays in.

$$\text{Revenue} = \beta_0 + \beta_1 \,\text{Wins} + \beta_2 \,\text{Enrollment} + \sum_c \delta_c \, D + \varepsilon \tag{4}$$

In this case after performing this regression we can then utilize the beta 1 in place of Marginal Revenue as the dollar value to a Power 4 team added for winning one extra game in the regular season. This gives us our final variable needed to estimate a Quarterback's value added to a P4 college football team.

Variable	Estimate (USD)
Regular Season Wins	4,192,000
Enrollment	2,042
ACC	18,130,000
Big 12	7,582,000
Big Ten	1,510,000
SEC	27,640,000

This gives us our final framework for estimating a quarterback's value as the following:

$$\text{MRP} = 4.2Million * 2.559Y(1-Y)\left(\frac{dPF}{PF} - \frac{dPA}{PA}\right) \tag{5}$$

Conclusion & Discussion

The unique nature of the current state of college football lends itself quite handily to this sort of method for evaluating a single player's value because it gives a framework to quickly establish single year contracts. Although the players are not technically considered employees of the university, they are however functioning as employees. They generate revenue for the school, and are now sharing in that revenue.

Table 1: Top 5 Players by MRP

Player	Team	MRP (USD)
Taylen Green	Arkansas	1,156,416
Brendan Sorsby	Cincinnati	1,125,484
Josh Hoover	TCU	1,076,855
Luke Altmyer	Illinois	1,068,626
Darian Mensah	Duke	1,058,406

When running our final equation on this season's data we see some interesting results emerge. In the top 5 players by their marginal revenue produced we don't see any of the Heisman finalists, but rather quarterbacks from institutions in tough conferences that have significantly overperformed their recruiting position. This would make sense as the quarterback would then have an outsized influence on the team's performance. However one of the most interesting findings is that our top performers are only estimated to be worth around a million dollars, and most players are worth quite a bit less. This would fly in the face of some of the reported values for NIL deals that certain players are signing. This is not a discouraging observation however, as we believe the current market to be inefficient, and also the players are supposedly profiting off of revenue generated off the field for outside institutions. Whether that is actually the truth is a question outside the purview of this exercise.

Despite the promising outcome there are some limitations to be discussed. The final regression for marginal revenue created by a football win resulted in an r squared value around 0.4, meaning it only explained around 40% of the overall variance in the final revenues. While this is a limitation while looking from the outside, this would offer an opportunity for an individual school to utilize datapoints unique to their school to find a marginal revenue value unique to their situation.

One other opportunity offered by this framework for a school attempting to use it themselves is the estimate for an individual's impact on a team's scoring. While I do believe the Predicted Points Added statistic is one of the better public estimates available for an individual's impact on team scoring, each team will have tracking data, film, and grading statistics available to them that are not publicly available that would be better used for a team attempting to apply this method to their own roster.

Finally, the biggest limitation when attempting to estimate the marginal revenue portion of this equation is still a lack of information. In today's day and age team's receive a significant portion of the money available to them from donor contribution's to outside companies for NIL purposes and this data would not be available in the government reports on revenue. This also applies to private schools like Baylor, Notre Dame, and Northwestern who have the ability to opt out of reports like this. It would be most important for a team attempting to apply this framework to their own roster to not explicitly accept the numbers calculated here as fact, but rather the methods and structure as sound to apply to their unique situation.

Reference and Sources

"A dynamic pythagorean exponent: Applications to baseball and structural modeling in risk management" – Albert Cohen

Knight-Newhouse College Athletics Database - https://knightnewhousedata.org/reports

Integrated Postsecondary Education Data System - https://nces.ed.gov/ipeds